

18.11.03

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

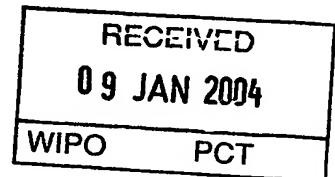
This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application: 2003年 7月31日

出願番号
Application Number: 特願2003-284489

[ST. 10/C]: [JP2003-284489]

出願人
Applicant(s): 松下電器産業株式会社

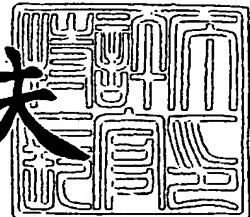


PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2003年12月22日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



【書類名】 特許願
【整理番号】 2033850173
【あて先】 特許庁長官殿
【国際特許分類】 G10L
【発明者】
【住所又は居所】 大阪府門真市大字門真1006番地 松下電器産業株式会社内
【氏名】 芳澤 伸一
【特許出願人】
【識別番号】 000005821
【氏名又は名称】 松下電器産業株式会社
【代理人】
【識別番号】 100109210
【弁理士】
【氏名又は名称】 新居 広守
【先の出願に基づく優先権主張】
【出願番号】 特願2003- 89179
【出願日】 平成15年 3月27日
【手数料の表示】
【予納台帳番号】 049515
【納付金額】 21,000円
【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【包括委任状番号】 0213583

【書類名】特許請求の範囲**【請求項 1】**

音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、一定の属性を有する音声の特徴を示す確率モデルである1以上の参照モデルを記憶するための参照モデル記憶手段と、

外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、及び、新たな参照モデルを作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段と、

所定の構造をもつ当該標準モデルの統計量の初期値を準備し、前記参照モデル記憶手段に格納された1以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するよう、前記参照モデルの統計量を用いて当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段と

を備えることを特徴とする音声認識用の標準モデル作成装置。

【請求項 2】

前記標準モデル作成装置はさらに、

音声認識の対象となる属性に関する情報である利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から1以上の参照モデルを選択する参照モデル選択手段を備え、

前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルに対する前記標準モデルの確率又は尤度を最大化又は極大化するよう前記標準モデルの統計量を計算する

ことを特徴とする請求項1記載の音声認識用の標準モデル作成装置。

【請求項 3】

音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、

一定の属性を有する音声の特徴を示す確率モデルである1以上の参照モデルを記憶する参照モデル記憶手段と、

音声認識の対象となる属性に関する情報である利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から1以上の参照モデルを選択する参照モデル選択手段と、

所定の構造をもつ当該標準モデルの統計量の初期値を準備し、前記参照モデル選択手段が選択した1以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するよう、前記参照モデルの統計量を用いて当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段と

を備えることを特徴とする音声認識用の標準モデル作成装置。

【請求項 4】

前記標準モデル作成手段はさらに、

当該標準モデルの構造を決定する標準モデル構造決定部を備え、

前記標準モデル構造決定部は、作成する標準モデルの仕様に関する情報である仕様情報、及び、前記利用情報の少なくともいずれか一方に基づいて、当該標準モデルの構造を決定し、

前記標準モデル作成手段は、前記標準モデル構造決定部が決定した構造をもつ当該標準モデルの統計量の初期値を準備し、前記参照モデル記憶手段に格納された1以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するよう、前記参照モデルの統計量を用いて当該標準モデルの統計量を計算することによって標準モデルを作成することによって標準モデルを作成することを特徴とする請求項1～請求項3のいずれか1項に記載の音声認識用の標準モデル作成装置。

【請求項 5】

音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の

属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、一定の属性を有する音声の特徴を示す確率モデルである1以上の参照モデルを記憶するための参照モデル記憶手段と、

所定の構造をもつ当該標準モデルの統計量の初期値を準備し、前記参照モデル記憶手段に格納された1以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するように、前記参照モデルの統計量を用いて当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備え、

前記標準モデル作成手段は、

当該標準モデルの構造を決定する標準モデル構造決定部を備え、

前記標準モデル構造決定部は、作成する標準モデルの仕様に関する情報である仕様情報、及び、音声認識の対象となる属性に関する情報である利用情報の少なくともいずれか一方に基いて、当該標準モデルの構造を決定し、

前記標準モデル作成手段は、前記標準モデル構造決定部が決定した構造をもつ当該標準モデルの統計量の初期値を準備する。

ことを特徴とする音声認識用の標準モデル作成装置。

【請求項6】

前記仕様情報とは、標準モデルを使用するアプリケーションプログラムの種類、及び、標準モデルを使用する機器の仕様の少なくともいずれか一方の仕様を示す

ことを特徴とする請求項4又は請求項5に記載の音声認識用の標準モデル作成装置。

【請求項7】

前記属性とは、年齢、性別、声質の話者性、感情又は健康状態に基づく声の調子、発話速度、発話の丁寧さ、方言、背景雑音の種類、背景雑音の大きさ、音声と背景雑音とのSN比、マイク特性、の少なくとも1つに関する情報を含む

ことを特徴とする請求項1～請求項6のいずれか1項に記載の音声認識用の標準モデル作成装置。

【書類名】明細書

【発明の名称】標準モデル作成装置および標準モデル作成方法

【技術分野】

【0001】

本発明は、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ベイジアンネットなどの確率モデルによる意図理解（意図の認識）、確率モデルによるデータマイニング（データ特性の認識）、確率モデルによる人物検出、指紋認証、顔認証、虹彩認証（対象を認識して特定の対象かどうかを判断する）、株価予測、天気予測などの予測（状況を認識して判断する）などに用いられる標準モデルの作成装置およびその方法に関する。

【背景技術】

【0002】

近年、インターネットなどの普及により、ネットワークの大容量化、通信コストの低価格化が進んでいる。このため、ネットワークを利用することで、多くの認識用モデル（参照モデル）を収集することが可能となってきている。例えば、音声認識において、様々な研究機関で配布している多くの音声認識用モデル（子供用モデル、成人用モデル、高齢者用モデル、自動車内用モデル、携帯電話用モデルなど）をインターネットによりダウンロードすることが可能となってきている。また、ネットワークによる機器連携により、カーナビゲーションシステムなどで利用する音声認識用モデルをテレビやパソコンなどにダウンロードできるようになってきている。また、意図理解において、各地の様々な人の経験を学習した認識用モデルを、ネットワークを通して収集することが可能となってきている。

【0003】

また、認識技術の発展により、認識用モデルは、パソコン、テレビのリモコン、携帯電話、カーナビゲーションシステムなど、CPUパワー、メモリ量などの仕様の異なる幅広い機器に利用されるようになってきている。また、セキュリティーなどの認識精度が要求されるアプリケーションや、テレビのリモコンでの操作のように認識結果が出力されるまでの時間の速さが要求されるアプリケーションなど、要求仕様の異なる幅広いアプリケーションに利用されるようになってきている。

【0004】

また、認識技術は、認識対象の異なる多くの環境で利用されるようになってきている。例えば、音声認識において、子供の声、成人の声、高齢者の声を認識したり、自動車内の声、携帯電話での声を認識するなど、多くの環境で利用される。

【0005】

これらの社会環境の変化を鑑みると、多くの認識用モデル（参照モデル）を有効に活用することで、機器やアプリケーションの仕様、利用環境に適した精度の高い認識用モデル（標準モデル）を短時間に作成して利用者に提供することが望まれると考えられる。

【0006】

音声認識などのパターン認識の分野では、認識用の標準モデルとして確率モデルを用いる方法が近年注目されており、特に、隠れマルコフモデル（以下HMMと呼ぶ）や混合ガウス分布モデル（以下GMMと呼ぶ）が広く用いられている。また、意図理解において、意図、知識、嗜好などを表す標準モデルとして確率モデルを用いる方法が近年注目されており、特に、ベイジアンネットなどが広く用いられている。また、データマイニングの分野で、データを分類するために各カテゴリの代表モデルとして確率モデルを用いる方法が注目されており、GMMなどが広く用いられている。また、音声認証、指紋認証、顔認証、虹彩認証などの認証の分野で、認証用の標準モデルとして確率モデルを用いる方法が注目されており、GMMなどが用いられている。HMMにより表現される標準モデルの学習アルゴリズムとしてバウム・ウェルチ（Baum-Welch）の再推定の方法が広く用いられている（例えば、今井聖著、『音声認識』、pp.150-152、共立出版株式会社、1995年11月25日発行参照）。また、GMMにより表現される標準モデルの学

習アルゴリズムとしてEM (Expectation-Maximization) アルゴリズムが広く用いられている（例えば、吉井貞▲ひろ▼著、”音声情報処理”、pp. 100-104、森北出版株式会社、1998年6月30日発行参照）。EMアルゴリズムでは、標準モデル

【0007】

【数1】

$$\sum_{m=1}^{M_f} \omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)$$

(ここで、

【0008】

【数2】

$$f(x; \mu_{f(m)}, \sigma_{f(m)}^2) \quad (m = 1, 2, \dots, M_f)$$

はガウス分布を表し、

【0009】

【数3】

$$x = (x_{(1)}, x_{(2)}, \dots, x_{(J)}) \in R^J$$

はJ (≥ 1) 次元の入力データを表す）における統計量である混合重み係数

【0010】

【数4】

$$\omega_{f(m)} \quad (m = 1, 2, \dots, M_f)$$

、J (≥ 1) 次元の平均値

【0011】

【数5】

$$\mu_{f(m)} = (\mu_{f(m,1)}, \mu_{f(m,2)}, \dots, \mu_{f(m,J)}) \in R^J$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

およびJ (≥ 1) 次元の分散値（共分散行列のJ個の対角成分）

【0012】

【数6】

$$\sigma_{f(m)}^2 = (\sigma_{f(m,1)}^2, \sigma_{f(m,2)}^2, \dots, \sigma_{f(m,J)}^2) \in R^J$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

を、

N個の学習データ

【0013】

【数7】

$$x[i] = (x_{(1)}[i], x_{(2)}[i], \dots, x_{(J)}[i]) \in R^J \quad (i=1,2,\dots,N)$$

を用いて、学習データに対する尤度

【0014】

【数8】

$$\log P = \sum_{i=1}^N \log \left[\sum_{m=1}^{M_f} \omega_{f(m)} f(x[i]; \mu_{f(m)}, \sigma_{f(m)}^2) \right]$$

を最大化もしくは極大化するように、

【0015】

【数9】

$$\omega_{f(m)} = \frac{\sum_{i=1}^N \gamma(x[i], m)}{\sum_{k=1}^{M_f} \sum_{i=1}^N \gamma(x[i], k)}$$

$$(m=1,2,\dots,M_f)$$

【0016】

【数10】

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^N \gamma(x[i], m) x_{(j)}}{\sum_{i=1}^N \gamma(x[i], m)}$$

$$(m=1,2,\dots,M_f, j=1,2,\dots,J)$$

【0017】

【数11】

$$\sigma_{f(m,j)}^2 = \frac{\sum_{i=1}^N \gamma(x[i], m) (x_{(j)} - \mu_{f(m,j)})^2}{\sum_{i=1}^N \gamma(x[i], m)}$$

$$(m=1,2,\dots,M_f, j=1,2,\dots,J)$$

(ここで、

【0018】

【数12】

$$\gamma(x[i], m) = \frac{\omega_{f(m)} f(x[i]; \mu_{f(m)}, \sigma_{f(m)}^2)}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x[i]; \mu_{f(k)}, \sigma_{f(k)}^2)} \quad (m = 1, 2, \dots, M_f)$$

である）を利用して1以上繰り返して計算して学習を行う。また、ベイズ推定法（例えば、繁樹算男著、”ベイズ統計入門”、pp.42-53、東京大学出版会、1985年4月30日発行参照）などの方法も提案されている。バウム・ウェルチの再推定の方法、EMアルゴリズム、ベイズ推定法のいずれの学習アルゴリズムも、学習データに対する確率（尤度）を最大化もしくは極大化するように標準モデルのパラメータ（統計量）を計算して標準モデルを作成する。これらの学習方法では、確率（尤度）を最大化もしくは極大化するという数学的な最適化が実現されている。

【0019】

上記の学習方法を音声認識の標準モデルの作成に用いた場合、多様な話者や雑音などの音響的特徴量の変動に対応するために多数の音声データで標準モデルを学習することが望ましい。また、意図理解に用いた場合、多様な話者や状況などの変動に対応するために多数のデータで標準モデルを学習することが望ましい。また、虹彩認証に用いた場合、太陽光、カメラ位置・回転などの変動に対応するために多数の虹彩画像データで標準モデルを学習することが望ましい。しかしながら、このような多量のデータを取り扱う場合、学習に膨大な時間がかかるため、利用者に標準モデルを短時間に提供できない。また、多量のデータを蓄積するためのコストが膨大となる。また、ネットワークを利用してデータを収集した場合、通信コストが膨大となる。

【0020】

一方、複数のモデル（以下、標準モデルの作成のために参照用として準備されるモデルを「参照モデル」と呼ぶ。）を合成することで標準モデルを作成する方法が提案されている。参照モデルは、多くの学習データを確率分布の母数（平均、分散など）で表現した確率分布モデルであり、多くの学習データの特徴を少数のパラメータ（母数）で集約したものである。以下に示す従来技術では、モデルはガウス分布で表現されている。

【0021】

第1の従来方法では、参照モデルはGMMで表現されており、複数の参照モデルのGMMを重み付きで合成することで標準モデルを作成している（例えば、特許文献1参照）。

【0022】

また、第2の従来方法では、第1の従来方式に加えて、学習データに対する確率（尤度）を最大化あるいは極大化して線形結合された混合重みを学習することで標準モデルを作成している（例えば、特許文献2参照）。

【0023】

また、第3の従来方法では、標準モデルの平均値を参照モデルの平均値の線形結合で表現し、入力データに対する確率（尤度）を最大化あるいは極大化して線形結合係数を学習することで標準モデルを作成している。ここでは学習データとして特定話者の音声データを用いており標準モデルを音声認識用の話者適応モデルとして用いている（例えば、非特許文献1参照）。

【0024】

また、第4の従来方法では、参照モデルは単一ガウス分布で表現されており、複数の参照モデルのガウス分布を合成したのちに、クラスタリングにより同一クラスに属するガウス分布を統合することで標準モデルを作成している（例えば、特許文献3参照）。

【0025】

また、第5の従来方法では、複数の参照モデルは同数の混合数の混合ガウス分布で表現され、各ガウス分布には1対1に対応した通し番号が付与されている。標準モデルは、同

一の通し番号をもつガウス分布を合成することにより作成される。合成する複数の参照モデルは利用者に音響的に近い話者で作成されたモデルであり、作成させる標準モデルは話者適応モデルである（例えば、非特許文献2参照）。

【特許文献1】特開平4-125599号公報（第3項、第1図）

【特許文献2】特開平10-268893号公報（第3-6頁、第1図）

【特許文献3】特開平9-81178号公報（第3-4項、第4図、第5図）

【非特許文献1】M. J. F. Gales, "Cluster Adaptive Training For Speech Recognition", 1998年、ICSLP'98予稿集、pp. 1783-1786

【非特許文献2】芳澤、外6名、"十分統計量と話者距離を用いた音韻モデルの教師なし学習法"、2002年3月1日、電子情報通信学会、Vol. J85-D-II、No. 3、pp. 382-389

【発明の開示】

【発明が解決しようとする課題】

【0026】

しかしながら、第1の従来方法では、合成する参照モデル数の増加とともに標準モデルの混合数が増加して、標準モデルのための記憶容量、認識処理量が膨大となり実用的でない。また、仕様に応じて標準モデルの混合数を制御することができない。この課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

【0027】

第2の従来方法では、合成する参照モデル数の増加とともに標準モデルの混合数が増加して、標準モデルのための記憶容量、認識処理量が膨大となり実用的でない。また、仕様に応じて標準モデルの混合数を制御することができない。また、標準モデルは、参照モデルの単純な混合和であり学習するパラメータが混合重みに限定されているため、高精度の標準モデルが作成できない。また、標準モデルの作成において、多くの学習データを用いて学習を行っているため学習時間がかかる。これらの課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

【0028】

第3の従来方法では、学習するパラメータが参照モデルの平均値の線形結合係数に限定されているため高精度の標準モデルが作成できない。また、標準モデルの作成において、多くの学習データを用いて学習を行っているため学習時間がかかる。

【0029】

第4の従来方法では、クラスタリングをヒューリスティックに行うため高精度の標準モデルを作成することが困難である。また、参照モデルは単一のガウス分布であるため精度が低く、それらを統合した標準モデルの精度は低い。認識精度に関する課題は、合成する参照モデルの数の増加に伴い顕著になってくると考えられる。

【0030】

第5の従来方法では、標準モデルは、同一の通し番号をもつガウス分布を合成することにより作成されるが、最適な標準モデルを作成するためには、一般的には合成するガウス分布は1対1に対応するとは限らないため、認識精度が低下する。また、複数の参照モデルが異なる混合数をもつ場合に標準モデルを作成することができない。また、一般的には、参照モデルにおけるガウス分布に通し番号が付与されておらず、この場合に標準モデルを作成することができない。また、仕様に応じて標準モデルの混合数を制御することができない。

【0031】

そこで、本発明は、このような問題点に鑑みてなされたものであり、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ペイジアンネットなどの確率モデルによる意図理解（意図の認識）、確率モデルによるデータマイニング（データ特性の認識）、株価予測、天気予測などの予測（状況を認識して判断する）などに用いられる高精度な標準モデルを作成する標準モ

ル作成装置等を提供することを目的とする。

【0032】

また、本発明は、学習のためのデータや教師データを必要とすることなく、簡易に標準モデルを作成することが可能な標準モデル作成装置等を提供することも目的とする。

【0033】

さらに、本発明は、標準モデルを利用する認識の対象にふさわしい標準モデルを作成したり、標準モデルを用いて認識処理を実行する装置の仕様や環境に適した標準モデルを作成することが可能な汎用性および柔軟性に優れた標準モデル作成装置等を提供することも目的とする。

【0034】

本発明で用いる「認識」とは、音声認識などの狭義の意味での認識だけではなく、パターンマッチング、識別、認証、ペイズ推定や予測など、確率で表現された標準モデルを利用するもの全般を意味する。

【課題を解決するための手段】

【0035】

上記目的を達成するために、本発明に係る標準モデル作成装置は、事象の集合の出力確率によって定義される認識用のモデルである標準モデル、あるいは、事象の集合と事象または事象間の遷移の出力確率とによって定義される認識用のモデルである標準モデルを作成する装置であって、特定の対象を認識するために予め作成されたモデルである参照モデルを1以上記憶する参照モデル記憶手段と、前記参照モデル記憶手段に記憶された1以上の参照モデルに対する標準モデルの確率または尤度を最大化または極大化するように当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備えることを特徴とする。

【0036】

これによって、1以上の参照モデルに対する標準モデルの確率または尤度を最大化または極大化するように標準モデルの統計量が計算され、標準モデルが作成されるので、音声データ等の学習データや教師データを必要とすることなく簡易に標準モデルが作成されるとともに、既に作成された複数の参照モデルを総合的に勘案した高精度な標準モデルが作成される。

【0037】

ここで、前記標準モデル作成装置は、さらに、外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、および、参照モデルを作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段を備えてもよい。例えば、音声認識用に適用した場合であれば、音声の特徴を示す周波数のパラメータを出力確率で表現する確率モデルを用いて、特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する装置であって、一定の属性を有する音声の特徴を示す確率モデルである1以上の参照モデルを記憶するための参照モデル記憶手段と、外部から参照モデルを取得して前記参照モデル記憶手段に格納すること、及び、新たな参照モデルを作成して前記参照モデル記憶手段に格納することの少なくとも一方を行う参照モデル準備手段と、所定の構造をもつ当該標準モデルの統計量の初期値を準備し、前記参照モデル記憶手段に格納された1以上の参照モデルに対する標準モデルの確率又は尤度を最大化又は極大化するように、前記参照モデルの統計量を用いて当該標準モデルの統計量を計算することによって標準モデルを作成する標準モデル作成手段とを備えることを特徴とする。

【0038】

これによって、標準モデル作成装置の外部から新たな参照モデルを取り込み、取り込んだ参照モデルに基づいた標準モデルの作成が可能となるので、様々な認識対象に対応した汎用性の高い標準モデル作成装置が実現される。

【0039】

また、前記標準モデル作成装置は、さらに、認識の対象に関する情報である利用情報を作成する利用情報作成手段と、作成された前記利用情報に基づいて、前記参照モデル記憶

手段に記憶されている参照モデルの中から1以上の参照モデルを選択する参照モデル選択手段とを備え、前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルに対する前記標準モデルの確率または尤度を最大化または極大化するように前記標準モデルの統計量を計算してもよい。

【0040】

これによって、利用者の特徴、利用者の年齢、性別、利用環境などの利用情報に基づいて、準備された複数の参照モデルの中から認識対象に適した参照モデルだけが選択され、それら参照モデルを統合した標準モデルが作成されるので、認識対象により特化した精度の高い標準モデルが作成される。

【0041】

ここで、前記標準モデル作成装置は、さらに、前記利用情報と選択された参照モデルに関する情報との類似度を算出して、前記類似度が所定のしきい値以上であるか否かを判定して判定信号を作成する類似度判定手段を備えてもよい。

これによって、利用情報にふさわしい（近い）参照モデルが参照モデル記憶手段に存在しない場合に、参照モデルの準備の要求を行うことができる。

【0042】

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、認識の対象に関する情報である利用情報を前記端末装置から受信する利用情報受信手段と、受信された前記利用情報に基づいて、前記参照モデル記憶手段に記憶されている参照モデルの中から1以上の参照モデルを選択する参照モデル選択手段とを備え、前記標準モデル作成手段は、前記参照モデル選択手段が選択した参照モデルに対する前記標準モデルの確率または尤度を最大化または極大化するように前記標準モデルの統計量を計算してもよい。

【0043】

これによって、通信路を介して送信されてきた利用情報に基づいて標準モデルが作成されるので、遠隔制御による標準モデルの生成が可能になるとともに、通信システムを基盤とする認識システムの構築が実現される。

【0044】

また、前記標準モデル作成装置は、さらに、作成する標準モデルの仕様に関する情報である仕様情報を作成する仕様情報作成手段を備え、前記標準モデル作成手段は、前記仕様情報作成手段が作成した仕様情報に基づいて、前記参照モデルに対する前記標準モデルの確率または尤度を最大化または極大化するように前記標準モデルの統計量を計算してもよい。

【0045】

これによって、標準モデルを使用する装置のCPUパワー、記憶容量、要求される認識精度、要求される認識処理時間などの仕様情報に基づいて標準モデルが作成されるので、特定の仕様条件を満たす標準モデルの生成が可能となり、計算エンジン等の認識処理に必要なリソース環境に適した標準モデルの生成が実現される。

【0046】

ここで、前記仕様情報は、例えば、標準モデルを使用するアプリケーションプログラムの種類に対応づけられた仕様を示すような情報であってもよい。そして、前記標準モデル作成装置は、さらに、標準モデルを使用するアプリケーションプログラムと標準モデルの仕様との対応を示すアプリケーション仕様対応データベースを前記仕様情報として保持する仕様情報保持手段を備え、前記標準モデル作成手段は、前記仕様情報保持手段に保持されたアプリケーション仕様対応データベースから、起動されるアプリケーションプログラムに対応する仕様を読み出し、読み出した仕様に基づいて、前記参照モデルに対する前記標準モデルの確率または尤度を最大化または極大化するように前記標準モデルの統計量を計算してもよい。

【0047】

これによって、各アプリケーションごとに對応づけられた仕様に沿って標準モデルが作

成されるので、アプリケーションごとに最適な標準モデルが作成され、標準モデルが使用される認識システム等における認識精度が向上される。

【0048】

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、作成する標準モデルの仕様に関する情報である仕様情報を前記端末装置から受信する仕様情報受信手段を備え、前記標準モデル作成手段は、前記仕様情報受信手段が受信した仕様情報に基づいて、前記参照モデルに対する前記標準モデルの確率または尤度を最大化または極大化するように前記標準モデルの統計量を計算してもよい。

【0049】

これによって、通信路を介して送信されてきた仕様情報に基づいて標準モデルが作成されるので、遠隔制御による標準モデルの生成が可能になるとともに、通信システムを基盤とする認識システムの構築が実現される。

【0050】

たとえば、前記参照モデルおよび前記標準モデルは、1以上のガウス分布を用いて表現され、前記標準モデル作成手段は、前記仕様情報に基づいて、前記標準モデルの混合分布数（ガウス分布の数）を決定してもよい。

【0051】

これによって、作成される標準モデルに含まれるガウス分布の混合分布数が動的に決定されることとなり、認識処理が実行される環境や要求仕様等に応じて標準モデルの構造を制御することが可能となる。例として、標準モデルを使用する認識装置のC P Uパワーが小さい場合、記憶容量が小さい場合、要求される認識処理時間が短い場合などは標準モデルの混合分布数を少なく設定して仕様に合わせることができ、一方、要求される認識精度が高い場合などは混合分布数を多く設定して認識精度を高くすることができる。

【0052】

なお、上記利用情報あるいは仕様情報を用いて標準モデルを作成する場合において、参照モデル準備手段は必ずしも必要ではない。たとえば、利用者の要求に基づいて、あるいは、利用者の要求とは無関係に、予め参照モデルを標準モデル作成装置内に記憶させた状態で標準モデル作成装置を出荷し、利用情報や仕様情報を用いて標準モデルを作成することが可能だからである。

【0053】

また、前記参照モデルおよび前記標準モデルは、1以上のガウス分布を用いて表現され、前記参照モデル記憶手段は、少なくとも1対の参照モデルの混合分布数（ガウス分布の数）が異なる参照モデルを記憶し、前記標準モデル作成手段は、少なくとも1対の参照モデルの混合分布数（ガウス分布の数）が異なる参照モデルに対する前記標準モデルの確率または尤度を最大化または極大化するように前記標準モデルの統計量を計算してもよい。

【0054】

これによって、混合分布数が異なる参照モデルに基づいて標準モデルが作成されるので、予め準備された多種多様な構造の参照モデルに基づく標準モデルの作成が可能となり、より認識対象に適した精度の高い標準モデルの作成が実現される。

【0055】

また、前記標準モデル作成装置は、さらに、前記標準モデル作成手段が作成した標準モデルを記憶する標準モデル記憶手段を備えてもよい。

これによって、作成された標準モデルを一時的にバッファリングしておき、送信要求に対してすぐに出力したり、他の装置に提供するデータサーバとしての役割を果たしたりすることが可能となる。

【0056】

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記標準モデル作成手段が作成した標準モデルを前記端末装置に送信する標準モデル送信手段を備えてもよい。

これによって、作成された標準モデルは空間的に離れた場所に置かれた外部装置に送信されるので、本標準モデル作成装置を標準モデル作成エンジンとして独立させたり、標準モデル作成装置を通信システムにおけるサーバとして機能させたりすることが可能になる。

【0057】

また、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記端末装置から送信される参照モデルを受信する参照モデル受信手段を備え、前記標準モデル作成手段は、少なくとも前記参照モデル受信手段が受信した参照モデルに対する前記標準モデルの確率または尤度を最大化または極大化するよう前記標準モデルの統計量を計算してもよい。

これによって、端末装置が保持した利用環境にふさわしい参照モデルを、通信路を介して送信して、送信した参照モデルを用いて標準モデルを作成できるため、より認識対象に適した精度の高い標準モデルの作成が実現される。例として、利用者Aが環境Aで利用していた参照モデルAが端末装置に保持されており利用者Aは環境Bで利用したい場合、参照モデルAを利用して標準モデルを作成することにより、利用者Aの特徴を反映した精度の高い標準モデルを作成することができる。

【0058】

また、前記参照モデル準備手段は、さらに、前記参照モデル記憶手段が記憶する参照モデルの更新および追加の少なくとも一方を行ってもよい。たとえば、前記標準モデル作成装置には、通信路を介して端末装置が接続され、前記標準モデル作成装置は、さらに、前記端末装置から送信される参照モデルを受信する参照モデル受信手段を備え、前記参照モデル準備手段は、前記参照モデル受信手段が受信した参照モデルを用いて前記参照モデル記憶手段が記憶する参照モデルの更新および追加の少なくとも一方を行ってもよい。

【0059】

これによって、準備される参照モデルの追加、更新等が行われるので、様々な認識対象用のモデルを参照モデルとして追加したり、より精度の高い参照モデルに置き換えたりすることが可能となり、更新した参照モデルによる標準モデルの再生成や、生成された標準モデルを参照モデルとして再び標準モデルを作成するというフィードバックによる学習等が可能となる。

【0060】

また、前記標準モデル作成手段は、作成する標準モデルの構造を決定する標準モデル構造決定部と、構造が決定された前記標準モデルを特定する統計量の初期値を決定する初期標準モデル作成部と、前記参照モデルに対する前記標準モデルの確率または尤度を最大化または極大化するように前記標準モデルの統計量を推定して計算する統計量推定部とを有するように構成してもよい。このとき、前記初期標準モデル作成部は、前記統計量推定部が標準モデルの統計量を計算するために用いる、1以上の前記参照モデルを用いて前記標準モデルを特定する統計量の初期値を決定してもよい。たとえば、前記初期標準モデル作成部は、標準モデルの種類を識別するクラスIDに基づいて、前記初期値を決定してもよい。具体的には、前記初期標準モデル作成部は、前記クラスIDと前記初期値と前記参照モデルとの対応を示す対応表を保持し、前記対応表に従って、前記初期値を決定してもよい。

【0061】

これによって、標準モデルが使用される認識の対象の種類ごとにクラスIDを付与しておくことで、最終的に必要とされる標準モデルと共通の性質をもつ初期標準モデルを使用することができるので、精度の高い標準モデルが作成される。

【0062】

なお、本発明は、このような標準モデル作成装置として実現することができるだけでなく、標準モデル作成装置が備える特徴的な構成要素をステップとする標準モデル作成方法として実現したり、それらのステップをコンピュータに実行させるプログラムとして実現したりすることができる。そして、そのプログラムをCD-ROM等の記録媒体やインターネット等の伝送媒体を介して配信することができるのは言うまでもない。

【発明の効果】**【0063】**

以上の説明から明らかなように、本発明に係る標準モデル作成装置によれば、1以上の参照モデルに対する標準モデルの確率または尤度を最大化または極大化するように標準モデルの統計量が計算され、標準モデルが作成されるので、学習データや教師データを必要とすることなく簡易に標準モデルが作成されるとともに、既に作成された複数の参照モデルを総合的に勘案した高精度な標準モデルが作成される。

【0064】

また、標準モデル作成装置の外部から新たな参照モデルを取り込み、取り込んだ参照モデルに基づいた標準モデルが作成されるので、様々な認識対象に対応した汎用性の高い標準モデル作成装置が実現される。

【0065】

また、利用者の特徴、利用者の年齢、性別、利用環境などの利用情報に基づいて選択された参照モデルを基準として、標準モデルが作成されるので、認識対象により特化した精度の高い標準モデルが作成される。

【0066】

また、通信路を介して送信されてきた参照モデルに基づいて標準モデルが作成されるので、遠隔制御による標準モデルの生成が可能になるとともに、通信システムを基盤とする認識システムの構築が実現される。

【0067】

また、標準モデルを使用する装置のC P Uパワー、記憶容量、要求される認識精度、要求される認識処理時間などの仕様情報に基づいて標準モデルが作成されるので、特定の仕様条件を満たす標準モデルの生成が可能となり、計算エンジン等の認識処理に必要なリソース環境に適した標準モデルの生成が実現される。

【0068】

また、通信路を介して送信されてきた仕様情報に基づいて標準モデルが作成されるので、遠隔制御による標準モデルの生成が可能になるとともに、通信システムを基盤とする認識システムの構築が実現される。

【0069】

また、仕様情報に基づいて、標準モデルに含まれるガウス分布の混合分布数が動的に決定されるので、認識処理が実行される環境や要求仕様等に応じて標準モデルの構造を制御することが可能となる。たとえば、標準モデルを使用する認識装置のC P Uパワーが小さい場合、記憶容量が小さい場合、要求される認識処理時間が短い場合などは標準モデルの混合数を少なく設定して仕様に合わせることができ、一方、要求される認識精度が高い場合は混合数を多く設定して認識精度を高くすることができる。

【0070】

また、仕様情報が各アプリケーションに対応づけられているので、アプリケーションごとに最適な標準モデルが作成され、標準モデルが使用される認識システム等における認識精度が向上される。

【0071】

また、混合分布数が異なる参照モデルに基づいて標準モデルが作成されるので、予め準備された多種多様な構造の参照モデルに基づく標準モデルの作成が可能となり、より認識対象に適した精度の高い標準モデルの作成が実現される。

【0072】

また、作成された標準モデルは内部の記憶手段に保持されるので、作成された標準モデルを一時的にバッファリングしておき、送信要求に対してすぐに出力したり、他の装置に提供するデータサーバとしての役割を果たしたりすることが可能となる。

【0073】

また、作成された標準モデルは空間的に離れた場所に置かれた外部装置に送信されるので、本標準モデル作成装置を標準モデル作成エンジンとして独立させたり、標準モデル作

成装置を通信システムにおけるサーバとして機能させたりすることが可能になる。

【0074】

また、準備される参照モデルの追加、更新等が行われるので、様々な認識対象用のモデルを参照モデルとして追加したり、より精度の高い参照モデルに置き換えたりすることが可能となり、更新した参照モデルによる標準モデルの再生成や、生成された標準モデルを参照モデルとして再び標準モデルを作成するというフィードバックによる学習等が可能となる。

【0075】

また、標準モデルが使用される認識の対象の種類ごとに付与されたクラスIDに基づいて初期標準モデルが決定され、その初期標準モデルに基づいて標準モデルが作成されるので、最終的に必要とされる標準モデルと共に性質をもつ初期標準モデルが使用され、短時間で精度の高い標準モデルが作成される。

【0076】

以上のように、本発明により、隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ペイジアンネットなどの確率モデルによる意図理解（意図の認識）、確率モデルによるデータマイニング（データ特性の認識）、確率モデルによる人物検出、指紋認証、顔認証、虹彩認証（対象を認識して特定の対象かどうかを判断する）、株価予測、天気予測などの予測（状況を認識して判断する）などに用いる高精度な標準モデルが提供され、その実用的価値は極めて高い。

【発明を実施するための最良の形態】

【0077】

以下、本発明の実施の形態について図面を参考しながら詳しく説明する。なお、図中同一または相当部分には同一符号を付し、その説明は繰り返さない。

(第1の実施の形態)

図1は、本発明の第1の実施の形態における標準モデル作成装置の全体構成を示すプロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ101に組み込まれた例が示されている。本実施の形態では特定の属性を有する音声の特徴を示す音声認識用の標準モデルを作成する場合を例にして説明する。

【0078】

サーバ101は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象または事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部111と、参照モデル準備部102と、参照モデル記憶部103と、標準モデル作成部104と、書き込み部112とを備える。

【0079】

読み込み部111は、CD-ROMなどのストレージデバイスに書き込まれた子供用参照モデル、成人用参照モデル、高齢者用参照モデルを読み込む。参照モデル準備部102は、読み込まれた参照モデル121を参照モデル記憶部103へ送信する。参照モデル記憶部103は、3個の参照モデル121を記憶する。ここで、参照モデルとは、標準モデルを作成するに際して参照される予め作成されたモデル（ここでは、音声認識用のモデル、つまり、一定の属性を有する音声の特徴を示す確率モデル）である。

【0080】

標準モデル作成部104は、参照モデル記憶部103が記憶した3個 ($Ng=3$) の参照モデル121に対する確率または尤度を最大化または極大化するように標準モデル122を作成する処理部であり、標準モデルの構造（ガウス分布の混合数など）を決定する標準モデル構造決定部104aと、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部104bと、決定された初期標準モデルを記憶する統計量記憶部104cと、統計量記憶部104cに記憶された初期標準モデルに対して、第1近似部104eによる近似計算等を用いることにより、参照モデル

記憶部103に記憶されている3個 ($N_g = 3$) の参照モデル121に対する確率または尤度を最大化または極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部104dとからなる。なお、統計量とは、標準モデルを特定するパラメータであり、ここでは、混合重み係数、平均値、分散値である。

【0081】

書き込み部112は、標準モデル作成部104が作成した標準モデル122をCD-ROMなどのストレージデバイスに書き込む。

【0082】

次に、以上のように構成されたサーバ101の動作について説明する。

図2は、サーバ101の動作手順を示すフローチャートである。

【0083】

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（ステップS100）。つまり、読み込み部111は、CD-ROMなどのストレージデバイスに書き込まれた子供用参照モデル、成人用参照モデル、高齢者用参照モデルを読み込み、参照モデル準備部102は、読み込まれた参照モデル121を参照モデル記憶部103へ送信し、参照モデル記憶部103は、3個の参照モデル121を記憶する。

【0084】

参照モデル121は、音素ごとのHMMにより構成される。参照モデル121の一例を図3に示す。ここでは、子供用参照モデル、成人用参照モデル、高齢者用参照モデルのイメージ図が示されている（なお、本図では、高齢者用参照モデルのイメージ図は省略されている）。これら3個の参照モデルの全てが、状態数3個、各状態は混合分布数が3個の混合ガウス分布によりHMMの出力分布が構成される。特徴量として12次元 ($J = 12$) のケプストラム係数が用いられる。

【0085】

次に、標準モデル作成部104は、参照モデル記憶部103が記憶した3個の参照モデル121に対する確率または尤度を最大化または極大化するように標準モデル122を作成する（ステップS101）。

【0086】

最後に、書き込み部112は、標準モデル作成部104が作成した標準モデル122をCD-ROMなどのストレージデバイスに書き込む（ステップS102）。CD-ROMなどのストレージデバイスに書き込まれた標準モデルは、子供、成人、高齢者を考慮した音声認識用の標準モデルとして利用される。

【0087】

図4は、図2におけるステップS101（標準モデルの作成）の詳細な手順を示すフローチャートである。

まず、標準モデル構造決定部104aは、標準モデルの構造を決定する（ステップS102a）。ここでは、標準モデルの構造として、音素ごとのHMMにより構成され、3状態であり、各状態における出力分布の混合数を3個 ($M_f = 3$) と決定する。

【0088】

次に、初期標準モデル作成部104bは、標準モデルを計算するための統計量の初期値を決定する（ステップS102b）。ここでは、参照モデル記憶部103に記憶された3つの参照モデルを、統計処理計算を用いて1つのガウス分布に統合したものを統計量の初期値とし、その初期値を初期標準モデルとして統計量記憶部104cに記憶する。

【0089】

具体的には、初期標準モデル作成部104bは、上記3つの状態I ($I = 1, 2, 3$) それぞれについて、以下の数13に示される出力分布を生成する。なお、式中のMf (ガウス分布の混合数) は、ここでは、3である。

【0090】

【数13】

$$\sum_{m=1}^{M_f} \omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)$$

ここで、

【0091】

【数14】

$$f(x; \mu_{f(m)}, \sigma_{f(m)}^2) \quad (m = 1, 2, \dots, M_f)$$

は、ガウス分布を表し、

【0092】

【数15】

$$x = (x_{(1)}, x_{(2)}, \dots, x_{(J)}) \in R^J$$

は、12次元 ($J = 12$) のLPCケプストラム係数を表し、

【0093】

【数16】

$$\omega_{f(m)} \quad (m = 1, 2, \dots, M_f)$$

は、各ガウス分布の混合重み係数を表し、

【0094】

【数17】

$$\mu_{f(m)} = (\mu_{f(m,1)}, \mu_{f(m,2)}, \dots, \mu_{f(m,J)}) \in R^J \quad (m = 1, 2, \dots, M_f)$$

は、各ガウス分布の平均値を表し、

【0095】

【数18】

$$\sigma_{f(m)}^2 = (\sigma_{f(m,1)}^2, \sigma_{f(m,2)}^2, \dots, \sigma_{f(m,J)}^2) \in R^J \quad (m = 1, 2, \dots, M_f)$$

は、各ガウス分布の分散値を表す。

【0096】

そして、統計量推定部104dは、参照モデル記憶部103に記憶された3つの参照モデル121を用いて、統計量記憶部104cに記憶された標準モデルの統計量を推定する（ステップS102c）。

【0097】

具体的には、3つ ($N_g = 3$) の参照モデル121の各状態I ($I = 1, 2, 3$) における出力分布、即ち、以下の数19に示される出力分布に対する標準モデルの確率または尤度（以下の数25に示される尤度 $\log P$ ）を極大化もしくは最大化するような標準モデルの統計量（上記数16に示される混合重み係数、上記数17に示される平均値、および、上記数18に示される分散値）を推定する。

【0098】

【数19】

$$\sum_{l=1}^{L_{g(i)}} v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \quad (i = 1, 2, \dots, N_g)$$

ここで、

【0099】

【数20】

$$g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \quad (i = 1, 2, \dots, N_g, l = 1, 2, \dots, L_{(i)})$$

はガウス分布を表し、

【0100】

【数21】

$$L_{g(i)} \quad (i = 1, 2, \dots, N_g)$$

は各参照モデルの混合分布数（ここでは、3）を表し、

【0101】

【数22】

$$v_{g(i,l)} \quad (l = 1, 2, \dots, L_{g(i)})$$

は各ガウス分布の混合重み係数を表し、

【0102】

【数23】

$$\mu_{g(i,l)} \quad (l = 1, 2, \dots, L_{g(i)})$$

は各ガウス分布の平均値を表し、

【0103】

【数24】

$$\sigma_{g(i,l)}^2 \quad (l = 1, 2, \dots, L_{g(i)})$$

は各ガウス分布の分散値を表す。

【0104】

【数25】

$$\log P = \sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \log \left[\sum_{m=1}^{M_f} \omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2) \right] \left\{ \sum_{l=1}^{L_{g(i)}} v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx$$

そして、以下の数26、数27および数28に示される式に従って、それぞれ、標準モデルの混合重み係数、平均値および分散値を算出する。

【0105】

【数26】

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) \left\{ \sum_{l=1}^{L_{g(i)}} v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{k=1}^{M_f} \sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, k) \left\{ \sum_{l=1}^{L_{g(i)}} v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

$$(m = 1, 2, \dots, M_f)$$

【0106】

【数27】

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) x_{(j)} \left\{ \sum_{l=1}^{L_{g(i)}} v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) \left\{ \sum_{l=1}^{L_{g(i)}} v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

【0107】

【数28】

$$\sigma_{f(m,j)}^2 = \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) (x_{(j)} - \mu_{f(m,j)})^2 \left\{ \sum_{l=1}^{L_{g(i)}} v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \gamma(x, m) \left\{ \sum_{l=1}^{L_{g(i)}} v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

このとき、統計量推定部104dの第1近似部104eにより、以下の数29に示される近似式が用いられる。

【0108】

【数29】

$$\gamma(x, m) = \frac{\omega_{f(m)} f(x, \mu_{f(m)}, \sigma_{f(m)}^2)}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x, \mu_{f(k)}, \sigma_{f(k)}^2)} \approx \frac{\omega_{f(m)} f(x, \mu_{f(m)}, \sigma_{f(m)}^2)}{u_{h(m)} h(x, \mu_{h(m)}, \sigma_{h(m)}^2)}$$

$$(m = 1, 2, \dots, M_f)$$

ここで、

【0109】

【数30】

$$u_{h(m)} h(x; \mu_{h(m)}, \sigma_{h(m)}^2) \quad (m=1,2,\dots, M_f)$$

は、

【0110】

【数31】

$$u_{h(m)} \quad (m=1,2,\dots, M_f)$$

【0111】

【数32】

$$\mu_{h(m)} = (\mu_{h(m,1)}, \mu_{h(m,2)}, \dots, \mu_{h(m,J)}) \in R^J$$

を平均値とし、

【0112】

【数33】

$$\sigma_{h(m)}^2 = (\sigma_{h(m,1)}^2, \sigma_{h(m,2)}^2, \dots, \sigma_{h(m,J)}^2) \in R^J$$

を分散値とする単一のガウス分布を表す。

【0113】

また、第1近似部104eは、上記数30に示された単一ガウス分布の重み（数31）平均値（数32）および分散値（数33）を、それぞれ、以下の数34、数35および数36に示された式に従って算出する。

【0114】

【数34】

$$\omega_{f(m,p)} f(x; \mu_{f(m,p)}, \sigma_{f(m,p)}^2) \quad (m=1,2,\dots, M_f, p=1,2,\dots, P_{h(m)})$$

【0115】

【数35】

$$v_{g(i,l)} g(x; \mu_{g(i)}, \sigma_{g(i)}^2) \quad (i=1,2,\dots, N_g, l=1,2,\dots, L_{g(i)})$$

【0116】

【数36】

$$\omega_{f(k)} f(x; \mu_{f(k)}, \sigma_{f(k)}^2)$$

図5は、第1近似部104eによる近似計算を説明する図である。第1近似部104eは、本図に示されるように、上記数29に示された近似式における単一ガウス分布（数30）を、標準モデルを構成する全ての混合ガウス分布を用いて決定している。

【0117】

以上の第1近似部104eによる近似式を考慮してまとめると、統計量推定部104dでの計算式は次の通りになる。つまり、統計量推定部104dは、以下の数37、数38および数39に示される式に従って、それぞれ、混合重み係数、平均値および分散値を算出し、統計量記憶部104cに記憶する。そして、このような統計量の推定と統計量記憶部104cへの記憶をR(≥ 1)回、繰り返す。その結果得られた統計量を最終的に生成

する標準モデル122の統計量として出力する。

【0118】

【数37】

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \prod_{j=1}^J \sum_{l=1}^{L_{g(i)}} A_{(m,l,i,j)}}{\sum_{i=1}^{N_g} \sum_{k=1}^{M_f} \prod_{j=1}^J \sum_{l=1}^{L_{g(i)}} A_{(k,l,i,j)}} \quad (m=1,2,\dots,M_f)$$

$$A_{(m,l,j)} = \frac{\omega_{f(m)} \sigma_{g(l,j)} \sigma_{h(m,j)}^2}{\sqrt{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2}}$$

$$\times \exp \left\{ \frac{1}{2} \left[\left(\frac{\frac{\sigma_{f(m,j)} \sigma_{g(l,j)}}{\sigma_{h(m,j)}} \mu_{h(m,j)} - \frac{\sigma_{f(m,j)} \sigma_{h(m,j)}}{\sigma_{g(l,j)}} \mu_{g(l,j)} - \frac{\sigma_{g(l,j)} \sigma_{h(m,j)}}{\sigma_{f(m,j)}} \mu_{f(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2} \right)^2 + \frac{\mu_{h(m,j)} - \mu_{g(l,j)} - \mu_{f(m,j)}}{\sigma_{h(m,j)}^2 - \sigma_{g(l,j)}^2 - \sigma_{f(m,j)}^2} \right] \right\}$$

【0119】

【数38】

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_{g(i)}} B_{(m,l,i,j)}}{\sum_{i=1}^{N_g} \prod_{j=1}^J \sum_{l=1}^{L_{g(i)}} A_{(m,l,i,j)}} \quad (m=1,2,\dots,M_f, j=1,2,\dots,J)$$

$$B_{(m,l,i,j)} = \frac{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 \mu_{g(l,m,j)} + \sigma_{g(l,j)}^2 \sigma_{h(m,j)}^2 \mu_{f(m,j)} - \sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2 \mu_{h(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2} \times A_{(m,l,i,j)}$$

【0120】

【数39】

$$\sigma_{f(m,j)}^2 = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_{g(i)}} C_{(m,l,i,j)}}{\sum_{i=1}^{N_g} \prod_{j=1}^J \sum_{l=1}^{L_{g(i)}} A_{(m,l,i,j)}} \quad (m=1,2,\dots,M_f)$$

$$C_{(m,l,i,j)} = \left\{ \frac{\sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2 \sigma_{h(m,j)}^2}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2} \right.$$

$$\left. + \left(\mu_{f(m,j)} - \frac{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 \mu_{g(l,m,j)} + \sigma_{g(l,j)}^2 \sigma_{h(m,j)}^2 \mu_{f(m,j)} - \sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2 \mu_{h(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(l,j)}^2} \right)^2 \right\} \times A_{(m,l,i,j)}$$

なお、状態遷移確率については、HMMの対応する状態遷移確率を参照モデル121に對して全て加えあわせた全体が1になるように正規化したものを用いる。

【0121】

次に、本実施の形態をパーソナルコンピュータによる音声認識に適用した具体例を説明する。ここでは、サーバ101としてパソコン（PC）、読み込み部111としてCD-ROMドライブ装置を用いるものとし、標準モデルの具体的な使い方を中心に説明する。

【0122】

まず、利用者は、PC（サーバ101）のCD-ROMドライブ装置（読み込み部111）に、参照モデルとしての複数の音響モデルが格納された1枚のCD-ROMを装着する。そのCD-ROMには、例えば、「幼児」、「子供：男」、「子供：女」、「大人：男」、「大人：女」、「高齢者：男」、「高齢者：女」の各音響モデルが記憶されている。

【0123】

次に、利用者は、図6（a）及び（b）に示される画面表示例のように、PC（サーバ101）に接続されたディスプレイを用いて、家族構成（音声認識を利用する人）にあつた音響モデルを選択する。図6には、CD-ROMに記憶されている音響モデルが「CD-ROM」と書かれた枠内に表示され、それらの音響モデルの中から選択された音響モデルが「利用者」と書かれた枠内にコピーされる様子が示されている。ここでは、利用者の家族構成が、10歳の男の子と、50歳のお父さんと、40歳のお母さんの3人であるとし、利用者（お父さん）によって、「子供：男」、「大人：男」、「大人：女」の3個のモデルが「利用者」と書かれた枠内にドラッグして移動されている。このような操作によって、参照モデル準備部102による参照モデルの準備が行われる。つまり、3個の参照モデルが読み込み部111で読み出され、参照モデル準備部102を介して、参照モデル記憶部103に格納される。

【0124】

続いて、利用者は、図7（a）に示される画面表示例のように、作成する標準モデルの構造（混合分布数）を指定する。図7（a）では、「混合分布数」として「3個」、「10個」、「20個」が表示され、利用者は、これらの個数の中から希望するものを選択する。この操作によって、標準モデル構造決定部104aにより、これから作成する標準モデルの構造が決定される。

【0125】

なお、混合分布数の決定については、このような直接的な指定に限られず、例えば、図7（b）に示される画面表示例のように、利用者が選択した仕様情報に基づいて混合分布数を決定してもよい。図7（b）では、標準モデルを使用して音声認識を実行させる対象機器として、3種類の「利用機器」、つまり、「テレビ用」、「カーナビ用」、「携帯電話用」の中から利用機器を選択する様子が示されている。このとき、予め記憶された対応表に従って、例えば、「テレビ用」が選択された場合には混合分布数を3個と決定し、「カーナビ用」が選択された場合には混合分布数を20個と決定し、「携帯電話用」が選択された場合には混合分布数を10個と決定してもよい。

【0126】

その他、混合分布数の決定については、認識速度や精度、つまり、「素早く認識」、「通常」、「高精度に認識」の中から選択することで、それぞれの選択項目に対応した値（「素早く認識」=3個、「通常」=10個、「高精度に認識」=20個）を混合分布数として決定してもよい。

【0127】

このような入力操作が終了すると、初期標準モデル作成部104bによって初期標準モデルが作成された後に、統計量推定部104dによる繰り返し計算（学習）が行われ、標準モデルが作成される。このとき、図8の画面表示例に示されるように、標準モデル構造決定部104aによって、学習の進捗状況が表示される。利用者は、学習の進捗状況、学習終了時期などを知ることができ、標準モデルが完成されるまで安心して待つことができる。なお、進捗状況の表示として、例えば、図8（a）に示されるような学習度合いのバー表示、図8（b）に示されるような学習回数の表示、その他、尤度基準の表示等がある

。また、未学習時は一般的な顔画像を表示し、学習の完了に近づくにしたがって利用者の顔画像に変更していくような進捗表示であってもよい。同様に、未学習時には赤ちゃんを表示し、学習の完了に近づくにしたがって仙人を表示するような進捗表示であってもよい。

【0128】

このようにして標準モデルの作成が完了すると、作成された標準モデルは、標準モデル作成部104によってメモリカード（書き込み部112）に記録される。利用者は、そのメモリカードをPC（サーバ101の書き込み部112）から抜き出し、利用機器、例えば、テレビのメモリカード用スロットに挿入する。これによって、作成された標準モデルがPC（サーバ101）から利用機器（テレビ）に移動される。テレビは、装着されたメモリカードに記録された標準モデルを用いて、利用者（ここでは、テレビを利用する家族）を対象とした音声認識を行う。たとえば、テレビに付属したマイクに入力された音声を認識することによって、テレビ操作用のコマンドを判別し、そのコマンド（例えば、チャンネルの切り替え、EPGなどによる番組検索）を実行する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを用いた、音声によるテレビ操作が実現される。

【0129】

以上説明したように、本発明の第1の実施の形態によれば、予め準備された参照モデルに対する確率または尤度を最大化または極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、学習のためのデータや教師データを必要とすることなく簡単に標準モデルが作成されるとともに、既に作成された複数の参照モデルを総合的に勘案した精度の高い標準モデルが作成される。

【0130】

なお、標準モデル122は、音素ごとにHMMを構成するに限らず、文脈依存のHMMで構成してもよい。

また、標準モデル作成部104は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

【0131】

また、標準モデル122を構成するHMMは、音素ごとに異なる状態数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル121は、子供用参照モデル、成人用参照モデル、高齢者用参照モデルにおいて、異なる状態数により構成してもよいし、異なる混合数の混合ガウス分布により構成してもよい。

【0132】

また、標準モデル122を用いて、サーバ101において音声認識を行ってもよい。

また、参照モデル121をCD-ROM、DVD-RAMなどのストレージデバイスから読み込む代わりに、サーバ101において音声データから参照モデル121を作成してもよい。

【0133】

また、参照モデル準備部102は、必要に応じてCD-ROM、DVD-RAMなどのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部103に追加・更新してもよい。つまり、参照モデル準備部102は、新たな参照モデルを参照モデル記憶部103に格納するだけでなく、同一の認識対象についての参照モデルが参照モデル記憶部103に格納されている場合には、その参照モデルと置き換えることによって参照モデルを更新したり、参照モデル記憶部103に格納されている不要な参照モデルを削除してもよい。

【0134】

また、参照モデル準備部102は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部103に追加・更新してもよい。

【0135】

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。
 また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

【0136】

(第2の実施の形態)

図9は、本発明の第2の実施の形態における標準モデル作成装置の全体構成を示すプロック図である。ここでは、本発明に係る標準モデル作成装置がセットトップボックス201(以下、STBと呼ぶ)に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル(話者適応モデル)を作成する場合を例にして説明する。具体的には、STBによる音声認識機能により、テレビのEPG検索や番組切替、録画予約などを行う場合を例にして説明する。

【0137】

STB201は、ユーザの発話を認識してTV番組の自動切替等を行うデジタル放送用受信機であり、事象の集合と事象または事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、マイク211と、音声データ蓄積部212と、参照モデル準備部202と、参照モデル記憶部203と、利用情報作成部204と、参照モデル選択部205と、標準モデル作成部206と、音声認識部213とを備える。

【0138】

マイク211に収集された音声データは、音声データ蓄積部212に蓄積される。参照モデル準備部202は、音声データ蓄積部212が蓄積した音声データを用いて話者ごとに参照モデル221を作成し、参照モデル記憶部203に記憶する。

【0139】

利用情報作成部204は、利用情報224である利用者の音声をマイク211により収集する。ここで、利用情報とは、認識(狭義での認識、識別、認証など)の対象(人・物)に関する情報であり、ここでは、音声認識の対象となる利用者の音声である。参照モデル選択部205は、利用情報作成部204が作成した利用情報224に基づいて、参照モデル記憶部203が記憶している参照モデル221の中から、利用情報224が示す利用者の音声に音響的に近い参照モデル223を選択する。

【0140】

標準モデル作成部206は、参照モデル選択部205が選択した話者の参照モデル223に対する確率または尤度を最大化または極大化するように標準モデル222を作成する処理部であり、標準モデルの構造(ガウス分布の混合分布数など)を決定する標準モデル構造決定部206aと、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部206bと、決定された初期標準モデルを記憶する統計量記憶部206cと、統計量記憶部206cに記憶された初期標準モデルに対して、一般近似部206eによる近似計算等を用いることにより、参照モデル選択部205が選択した参照モデル223に対する確率または尤度を最大化または極大化するような統計量を算出する(最終的な標準モデルを生成する)統計量推定部206dとからなる。

【0141】

音声認識部213は、標準モデル作成部206によって作成された標準モデル222を用いて利用者の音声を認識する。

【0142】

次に、以上のように構成されたSTB201の動作について説明する。

図10は、STB201の動作手順を示すフローチャートである。

【0143】

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する(ステップS200)。つまり、マイク211によりAさんからZさんの音声データを収集して音声データ蓄積部212に蓄積する。たとえば、屋内に設置された複数のマイク、テレビのリモ

コンに内蔵されたマイク、電話機などが、STB201の音声データ蓄積部212と接続されており、マイクや電話機から入力された音声データを音声データ蓄積部212に蓄積する。たとえば、お兄ちゃん、妹、お父さん、お母さん、おじいちゃん、近所のひと、友達の音声が蓄積される。

【0144】

参照モデル準備部202は、音声データ蓄積部212が蓄積した音声データを用いて話者ごとに参照モデル221をバウム・ウェルチの再推定の方法により作成する。この処理は、標準モデルの作成が要求される以前に行われる。

【0145】

参照モデル記憶部203は、参照モデル準備部202が作成した参照モデル221を記憶する。参照モデル221は、音素ごとのHMMにより構成される。参照モデル221の一例を図11に示す。ここでは、AさんからZさんの全ての参照モデルが、状態数3個、各状態は混合分布数が5個の混合ガウス分布によりHMMの出力分布が構成される。特徴量として25次元($J=25$)のメルケプストラム係数が用いられる。

【0146】

ここで、標準モデルの作成が要求される。たとえば、利用者が「利用者の確認」のボタンを押すことによって、標準モデルの作成が要求される。「利用者確認」のボタンについては、テレビ画面に表示させて選択する方法や、テレビのリモコンに「利用者の確認」スイッチをつけて選択する方法が考えられる。ボタンを押すタイミングとしては、テレビを起動したタイミング、音声認識を用いてコマンド操作を行っているときに利用者にふさわしい標準モデルがほしいと感じたタイミングなどが考えられる。

【0147】

次に、利用情報作成部204は、利用情報224である利用者の音声をマイク211により収集する(ステップS201)。たとえば、標準モデルの作成が要求されると、画面上で「名前を入力してください」と表示される。利用者は、テレビのリモコンに内蔵されたマイクにより名前(利用者の音声)を入力する。この利用者の音声が利用情報である。なお、入力する音声は名前に限定されない。例えば「適応と発声してください」と表示して、利用者は「適応」と発声してもよい。

【0148】

参照モデル選択部205は、その利用者の音声に音響的に近い参照モデル223を、参照モデル記憶部203が記憶している参照モデル221の中から選択する(ステップS202)。具体的には、利用者の音声をAさんからZさんの参照モデルに入力して発声単語に対する尤度が大きい10人($Ng=10$)の話者の参照モデルを選択する。

【0149】

そして、標準モデル作成部206は、参照モデル選択部205が選択した10個の参照モデル223に対する確率または尤度を最大化または極大化するように標準モデル222を作成する(ステップS203)。このとき、第1の実施の形態のように、学習の進捗状況を表示してもよい。そうすることで、利用者は学習の進捗状況、学習終了時期などが判断でき、安心して標準モデルを作成することができる。また、学習の進捗状況を非表示にする進捗状況非表示部を設けてもよい。この機能により、画面を有効に使うことができる。また、慣れた人に対して非表示にすることで、うつとうしく感じることが回避される。

【0150】

最後に、音声認識部213は、マイク211から介して送られてくる利用者の音声を入力とし、標準モデル作成部206で作成された標準モデル222を用いて音声認識を行う(S204)。たとえば、利用者が発話した音声を音響解析等を行うことで25次元のメルケプストラム係数を算出し、音素ごとの標準モデル222に入力することで、高い尤度を有する音素の連なりを特定する。そして、その音素の連なりと予め受信している電子番組データ中の番組名とを比較し、一定以上の尤度が検出された場合に、その番組に切り替えるという自動番組切替の制御を行う。

【0151】

次に、図10におけるステップS203（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図4に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

【0152】

まず、標準モデル構造決定部206aは、標準モデルの構造を決定する（図4のステップS102a）。ここでは、標準モデルの構造として、音素ごとのHMMにより構成され、3状態であり、各状態における出力分布の混合分布数が16個（Mf=16）と決定する。

【0153】

次に、初期標準モデル作成部206bは、標準モデルを計算するための統計量の初期値を決定する（図4のステップS102b）。ここでは、参照モデル選択部205が選択した10個の参照モデル223を、統計処理計算を用いて1つのガウス分布に統合したものと統計量の初期値とし、その初期値を初期標準モデルとして統計量記憶部206cに記憶する。ここでは、話者ごとに学習した混合分布数が5の参照モデルを用いて精度の高い混合分布数が16（16混合）の標準モデル（話者適応モデル）を作成する。

【0154】

具体的には、初期標準モデル作成部206bは、上記3つの状態I（I=1, 2, 3）それぞれについて、上記数13に示される出力分布を生成する。

【0155】

ただし、本実施の形態では、上記数13に示された出力分布における

【数40】

$$x = (x_{(1)}, x_{(2)}, \dots, x_{(J)}) \in R^J$$

は、25次元（J=25）のメルケプストラム係数を表す。

【0156】

そして、統計量推定部206dは、参照モデル選択部205が選択した10個の参照モデル223を用いて、統計量記憶部206cに記憶された標準モデルの統計量を推定する（図4のステップS102c）。

【0157】

つまり、10個（Ng=10）の参照モデル223の各状態I（I=1, 2, 3）における出力分布、即ち、上記数19に示される出力分布に対する標準モデルの確率（ここでは、上記数25に示される尤度logP）を極大化もしくは最大化するような標準モデルの統計量（上記数16に示される混合重み係数、上記数17に示される平均値、および、上記数18に示される分散値）を推定する。

【0158】

ただし、本実施の形態では、上記数19に示された出力分布における

【数41】

$$L_{g(i)} \quad (i=1,2,\dots,N_g)$$

は、5（各参照モデルの混合分布数）である。

【0159】

具体的には、上記数26、数27および数28に示される式に従って、それぞれ、標準モデルの混合重み係数、平均値および分散値を算出する。

【0160】

このとき、統計量推定部206dの一般近似部206eにより、上記数29に示される近似式が用いられる。

【0161】

ここで、一般近似部206eは、第1の実施の形態と異なり、上記数29の近似式の分

母に示された出力分布
【数42】

$$\omega_{f(k)} f(x; \mu_{f(k)}, \sigma_{f(k)}^2) \quad (k=1,2,\dots, M_f)$$

の中から、上記数29の近似式の分子に示された出力分布
【数43】

$$\omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)$$

に距離的に近い3個 ($P_{h(m)} = 3$) の出力分布
【数44】

$$\omega_{f(m,p)} f(x; \mu_{f(m,p)}, \sigma_{f(m,p)}^2) \quad (m=1,2,\dots, M_f, p=1,2,\dots, P_{h(m)})$$

を選択し、選択した3個の出力分布を用いて、上記数30に示された単一ガウス分布の重み（数31）、平均値（数32）および分散値（数33）を、それぞれ、以下の数45、数46および数47に示された式に従って算出する。

【0162】
【数45】

$$u_{h(m)} = \sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)} \quad (m=1,2,\dots, M_f)$$

【0163】
【数46】

$$\mu_{h(m,j)} = \frac{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)} \mu_{f(m,p,j)}}{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)}} \quad (m=1,2,\dots, M_f, j=1,2,\dots, J)$$

【0164】
【数47】

$$\sigma_{h(m,j)}^2 = \frac{\sum_{p=1}^{P_{f(m)}} \omega_{f(m,p)} (\sigma_{f(m,p,j)}^2 + \mu_{f(m,p,j)}^2)}{\sum_{p=1}^{P_{h(m)}} \omega_{f(m,p)}} - \mu_{h(m,j)}^2$$

$$(m=1,2,\dots, M_f, j=1,2,\dots, J)$$

図12は、一般近似部206eによる近似計算を説明する図である。一般近似部206eは、本図に示されるように、上記数29に示された近似式における単一ガウス分布（数30）を、標準モデルを構成するMf個の混合ガウス分布の中から、計算対象となる混合ガウス分布に近い一部 ($P_{h(m)}$ 個) の混合ガウス分布だけを用いて決定している。したが

って、全部 (Mf個) の混合ガウス分布を用いる第1の実施の形態と比較し、近似計算における計算量が削減される。

【0165】

以上的一般近似部206eによる近似式を考慮してまとめると、統計量推定部206dでの計算式は次の通りになる。つまり、統計量推定部206dは、以下の数48、数49および数50に示される式に従って、それぞれ、混合重み係数、平均値および分散値を算出し、統計量記憶部206cに記憶する。そして、このような統計量の推定と統計量記憶部206cへの記憶をR (≥ 1) 回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル222の統計量として出力する。なお、繰り返し計算においては、その回数に対応させて、上記近似計算における出力分布の選択個数Ph(m)を小さくし、最終的に $Ph(m)=1$ とする計算を行う。

【0166】

【数48】

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(i)} \alpha_{(m,l,i)}}{\sum_{k=1}^{M_f} \omega_{f(k)} \left(\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(i)} \alpha_{(k,l,i)} \right)} \quad (m=1,2,\dots,M_f)$$

$$\alpha_{(m,l,i)} = v_{g(l,i)} \prod_{j=1}^J D_{(m,l,i,j)}$$

$$D_{(m,l,i,j)} = \frac{\sigma_{h(m,j)}^2}{\sqrt{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,j,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,j,j)}^2}}$$

$$\times \exp \left\{ \frac{1}{2} \left[\frac{\left(\frac{\sigma_{f(m,j)} \sigma_{g(i,j,j)}}{\sigma_{h(m,j)}} \mu_{h(m,j)} - \frac{\sigma_{f(m,j)} \sigma_{h(m,j)}}{\sigma_{g(i,j,j)}} \mu_{g(i,j,j)} - \frac{\sigma_{g(i,j,j)} \sigma_{h(m,j)}}{\sigma_{f(m,j)}} \mu_{f(m,j)} \right)^2}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,j,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,j,j)}^2} + \frac{\mu_{h(m,j)}}{\sigma_{h(m,j)}^2} \frac{\mu_{g(i,j,j)}}{\sigma_{g(i,j,j)}^2} \frac{\mu_{f(m,j)}}{\sigma_{f(m,j)}^2} \right] \right\}$$

【0167】

【数49】

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(i)} \beta_{(m,l,i,j)} \alpha_{(m,l,i)}}{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(i)} \alpha_{(m,l,i)}} \quad (m=1,2,\dots,M_f, j=1,2,\dots,J)$$

$$\beta_{(m,l,i,j)} = \frac{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 \mu_{g(i,j,j)} + \sigma_{g(i,j,j)}^2 \sigma_{h(m,j)}^2 \mu_{f(m,j)} - \sigma_{f(m,j)}^2 \sigma_{g(i,j,j)}^2 \mu_{h(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,j,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,j,j)}^2}$$

【0168】

【数50】

$$\sigma_{f(m,j)}^2 = \frac{\sum_{l=1}^{N_g} \sum_{i=1}^{L_g(l)} \gamma_{(m,l,i,j)} \alpha_{(m,l,i)}}{\sum_{i=1}^{N_g} \sum_{l=1}^{L_g(i)} \alpha_{(m,l,i)}} \quad (m=1,2,\dots,M_f, j=1,2,\dots,J)$$

$$\begin{aligned} \gamma_{(m,l,j)} &= \left\{ \frac{\sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2} \right. \\ &\quad \left. + \left(\mu_{f(m,j)} - \frac{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 \mu_{g(i,m,j)} + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 \mu_{f(m,j)} - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2 \mu_{h(m,j)}}{\sigma_{f(m,j)}^2 \sigma_{h(m,j)}^2 + \sigma_{g(i,l,j)}^2 \sigma_{h(m,j)}^2 - \sigma_{f(m,j)}^2 \sigma_{g(i,l,j)}^2} \right)^2 \right\} \end{aligned}$$

なお、状態遷移確率については、HMMの対応する状態遷移確率を参照モデル223に対して全て加えあわせた全体が1になるように正規化したものを用いる。

【0169】

以上説明したように、本発明の第2の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率または尤度を最大化または極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

【0170】

なお、標準モデルを作成するタイミングとしては、本実施の形態のような利用者による明示的な指示だけに限られず、他のタイミングで標準モデルを作成してもよい。たとえば、STB201にさらに、利用者が変更されたかどうかを自動的に判断する利用者変更判断部を設ける。その利用者変更判断部は、テレビのリモコンに入力された認識用の音声を用いて、利用者が変更されたか否か、つまり、現在の利用者が直前まで認識していた利用者と同一人物であるか否かを判断する。利用者が変更されたと判断した場合に、その音声を利用情報として標準モデルを作成する。これにより、利用者が意識することなく、利用者にふさわしい標準モデルを用いた音声認識が行われる。

【0171】

なお、標準モデル222は、音素ごとにHMMを構成するに限らず、文脈依存のHMMで構成してもよい。

また、標準モデル作成部206は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

【0172】

また、標準モデル222を構成するHMMは、音素ごとに異なる状態数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、参照モデル221は、話者ごとHMMにおいて、異なる状態数により構成してもよいし、異なる混合数の混合ガウス分布により構成してもよい。

【0173】

また、参照モデル221は、話者ごとHMMに限らず、話者・雑音・声の調子ごとに作成してもよい。

また、標準モデル222をCD-ROM、ハードディスク、DVD-RAMなどのストレージデバイスに記録してもよい。

【0174】

また、参照モデル221を作成する代わりに、CD-ROM、DVD-RAMなどのス

トレージデバイスから読み込んでもよい。

また、参照モデル選択部205は、利用情報224に基づいて利用者ごとに選択する参照モデルの数を変えてよい。

【0175】

また、参照モデル準備部202は、必要に応じて新たな参照モデルを作成して参照モデル記憶部203に追加・更新してもよいし、参照モデル記憶部203に格納されている不要な参照モデルを削除してもよい。

また、参照モデル準備部202は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部203に追加・更新してもよい。

【0176】

また、上記近似計算において選択する出力分布の個数Ph(m)は、対象とする事象や標準モデルの出力分布によって異なってもよいし、分布間距離に基づいて決定してもよい。

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

【0177】

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

また、混合分布数については、本実施の形態におけるSTBを出荷するときに、所定の値に設定しておいてもよいし、ネットワーク連携を考慮した機器のCPUパワーなどの仕様、起動するアプリケーションの仕様などに基づいて混合分布数を決定してもよい。

【0178】

(第3の実施の形態)

図13は、本発明の第3の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がPDA(Personal Digital Assistant)301に組み込まれた例が示されている。本実施の形態では雑音識別用の標準モデル(雑音モデル)を作成する場合を例にして説明する。

【0179】

PDA301は、携帯情報端末であり、事象の出力確率によって定義される雑音識別用の標準モデルを作成する標準モデル作成装置として、読み込み部311と、参照モデル準備部302と、参照モデル記憶部303と、利用情報作成部304と、参照モデル選択部305と、標準モデル作成部306と、仕様情報作成部307と、マイク312と、雑音識別部313とを備える。

【0180】

読み込み部311は、CD-ROMなどのストレージデバイスに書き込まれた乗用車Aの参照モデル、乗用車Bの参照モデル、バスAの参照モデル、小雨の参照モデル、大雨の参照モデルなどの雑音の参照モデルを読み込む。参照モデル準備部302は、読み込まれた参照モデル321を参照モデル記憶部303へ送信する。参照モデル記憶部303は、参照モデル321を記憶する。

【0181】

利用情報作成部304は、利用情報324である雑音の種類をPDA301の画面とキーを利用して作成する。参照モデル選択部305は、利用情報324である雑音の種類に音響的に近い参照モデルを、参照モデル記憶部303が記憶している参照モデル321の中から選択する。仕様情報作成部307は、PDA301の仕様に基づき仕様情報325を作成する。ここで、仕様情報とは、作成する標準モデルの仕様に関する情報であり、ここでは、PDA301が備えるCPUの処理能力に関する情報である。

【0182】

標準モデル作成部306は、仕様情報作成部307で作成された仕様情報325に基づいて、参照モデル選択部305が選択した雑音の参照モデル323に対する確率または尤度を最大化または極大化するように標準モデル322を作成する処理部であり、標準モデルの構造(ガウス分布の混合分布数など)を決定する標準モデル構造決定部306aと、

標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部306bと、決定された初期標準モデルを記憶する統計量記憶部306cと、統計量記憶部306cに記憶された初期標準モデルに対して、第2近似部306eによる近似計算等を用いることにより、参照モデル選択部305が選択した参照モデル323に対する確率または尤度を最大化または極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部306dとからなる。

【0183】

雑音識別部313は、標準モデル作成部306で作成された標準モデル322を用いて、マイク312から入力された雑音の種類を識別する。

【0184】

次に、以上のように構成されたPDA301の動作について説明する。

図14は、PDA301の動作手順を示すフローチャートである。

【0185】

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（ステップS300）。つまり、読み込み部311は、ストレージデバイスに書き込まれた雑音の参照モデルを読み込み、参照モデル準備部302は、読み込まれた参照モデル321を参照モデル記憶部303へ送信し、参照モデル記憶部303は、参照モデル321を記憶する。

【0186】

参照モデル321は、GMMより構成される。参照モデル321の一例を図15に示す。ここでは、各雑音モデルは混合分布数が3個のGMMにより構成される。特徴量として5次元（J=5）のLPCケプストラム係数が用いられる。

【0187】

次に、利用情報作成部304は、識別したい雑音の種類である利用情報324を作成する（ステップS301）。図16にPDA301の選択画面の一例を示す。ここでは、乗用車の雑音が選択される。参照モデル選択部305は、選択された利用情報324である乗用車の雑音に音響的に近い参照モデルである乗用車Aの参照モデルと乗用車Bの参照モデルを、参照モデル記憶部303が記憶している参照モデル321の中から選択する（ステップS302）。

【0188】

そして、仕様情報作成部307は、PDA301の仕様に基づき、仕様情報325を作成する（ステップS303）。ここでは、PDA301のCPUの仕様に基づきCPUパワーが小さいという仕様情報325を作成する。標準モデル作成部306は、作成された仕様情報325に基づいて、参照モデル選択部305が選択した参照モデル323に対する確率または尤度を最大化または極大化するように標準モデル322を作成する（ステップS304）。

【0189】

最後に、雑音識別部313は、利用者によってマイク312から入力された雑音に対して、標準モデル322を用いて、雑音の識別を行う（ステップS305）。

【0190】

次に、図14におけるステップS304（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図4に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

【0191】

まず、標準モデル構造決定部306aは、標準モデルの構造を決定する（図4のステップS102a）。ここでは、標準モデルの構造として、仕様情報325であるCPUパワーが小さいという情報に基づいて1混合（Mf=1）のGMMにより標準モデル322を構成すると決定する。

【0192】

次に、初期標準モデル作成部306bは、標準モデルを計算するための統計量の初期値を決定する（図4のステップS102b）。ここでは、選択された参照モデル323であ

る乗用車Aの3混合の参照モデルを、統計処理計算を用いて1つのガウス分布に統合したものを統計量の初期値として統計量記憶部306cに記憶する。

【0193】

具体的には、初期標準モデル作成部306bは、上記数13に示される出力分布を生成する。

【0194】

ただし、本実施の形態では、上記数13に示された出力分布における
【数51】

$$x = (x_{(1)}, x_{(2)}, \dots, x_{(J)}) \in R^J$$

は、5次元 ($J = 5$) のLPCケプストラム係数を表す。

【0195】

そして、統計量推定部306dは、参照モデル選択部305が選択した2個の参照モデル323を用いて、統計量記憶部306cに記憶された標準モデルの統計量を推定する（図4のステップS102c）。

【0196】

つまり、2個 ($N_g = 2$) の参照モデル323における出力分布、即ち、上記数19に示される出力分布に対する標準モデルの確率（ここでは、上記数25に示される尤度log P）を極大化もしくは最大化するような標準モデルの統計量（上記数16に示される混合重み係数、上記数17に示される平均値、および、上記数18に示される分散値）を推定する。

【0197】

ただし、本実施の形態では、上記数19に示された出力分布における
【数52】

$$L_{g(i)} \quad (i=1,2,\dots,N_g)$$

は、3（各参照モデルの混合分布数）である。

【0198】

具体的には、上記数26、数27および数28に示される式に従って、それぞれ、標準モデルの混合重み係数、平均値および分散値を算出する。

このとき、統計量推定部306dの第2近似部306eは、標準モデルの各ガウス分布はお互いに影響を与えない仮定して、以下の近似式を用いる。

【0199】

【数53】

$$\gamma(x, m) \approx \frac{\omega_{f(m)} f(x; \mu_{f(m)}, \sigma_{f(m)}^2)}{u_{h(m)} h(x; \mu_{h(m)}, \sigma_{h(m)}^2)} \approx 1.0$$

$$(m=1,2,\dots,M_f)$$

また、標準モデルのガウス分布

【数54】

$$\omega_{f(m,p)} f(x; \mu_{f(m,p)}, \sigma_{f(m,p)}^2) \quad (m=1,2,\dots,M_f, p=1,2,\dots,P_{h(m)})$$

の近傍の

【数55】

 x

とは、前記数54が示す出力分布との平均値のユークリッド距離、マハラノビス距離、カルバック・ライラー（KL）距離などの分布間距離が近いQg(m, i)個の参照モデル323のガウス分布

【数56】

$$g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \quad (i = 1, 2, \dots, N_g, l = 1, 2, \dots, L_{(i)})$$

が存在する空間であって、

【0200】

【数57】

$$\omega_{f(m,p)} f(x; \mu_{f(m,p)}, \sigma_{f(m,p)}^2) \quad (m = 1, 2, \dots, M_f, p = 1, 2, \dots, P_{h(m)})$$

との分布間距離が近いQg(m, i)個 ($1 \leq Qg(m, i) \leq Lg(i)$) の前記参照ベクトルの出力分布とは、前記参照モデルの出力分布

【数58】

$$\nu_{g(i,l)} g(x; \mu_{g(i)}, \sigma_{g(i)}^2) \quad (i = 1, 2, \dots, N_g, l = 1, 2, \dots, L_{g(i)})$$

のうち分布間距離が1番近い（近傍指示パラメータG=1）前記標準モデルの出力分布が前記数52である前記参照ベクトルの出力分布であると近似する。

【0201】

図17は、この統計量推定部306dによる統計量の推定手順を示す概念図である。各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布mであるガウス分布を用いて統計量の推定を行うことが示されている。

【0202】

図18は、第2近似部306eによる近似計算を説明する図である。第2近似部306eは、本図に示されるように、各参照モデルの各ガウス分布に対して、距離が最も近い標準モデルのガウス分布mを決定することで、上記数53に示された近似式を用いている。

【0203】

以上の第2近似部306eによる近似式を考慮してまとめると、統計量推定部306dでの計算式は次の通りになる。つまり、統計量推定部306dは、以下の数59、数60および数61に示される式に従って、それぞれ、混合重み係数、平均値および分散値を算出し、それらのパラメータによって特定される標準モデルを最終的な標準モデル322として生成する。

【0204】

【数59】

$$\omega_{f(m)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} v_{g(i,l)}}{\sum_{k=1}^{M_f} \sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} v_{g(i,l)}}$$

$$(m = 1, 2, \dots, M_f)$$

(ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布であるガウス分布に関する和を意味する。)

【0205】

【数60】

$$\mu_{f(m,j)} = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} v_{g(i,l)} \mu_{g(i,l,j)}}{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} v_{g(i,l)}}$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

(ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布であるガウス分布に関する和を意味する。)

【0206】

【数61】

$$\sigma_{f(m,j)}^2 = \frac{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} v_{g(i,l)} (\sigma_{g(i,l)}^2 + \mu_{g(i,l,j)}^2)}{\sum_{i=1}^{N_g} \sum_{l=1}^{Q_{g(m,l)}} v_{g(i,l)}} - \mu_{f(m,j)}^2$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

(ここで、分母、分子の和は、各参照モデルの各ガウス分布に対して、平均値のユークリッド距離、マハラノビス距離などの分布間距離が最も近いものが標準モデルのガウス分布であるガウス分布に関する和を意味する。)

【0207】

ただし、

【数62】

$$\sum_{i=1}^{N_g} Q_{g(m,i)} = 0 \quad (m = 1, 2, \dots, M_f)$$

の場合において、

(第1の方法) 混合重み係数、平均値、分散値を更新しない。

(第2の方法) 混合重み係数の値をゼロにして、平均値、分散値を所定の値にする。

(第3の方法) 混合重み係数の値を所定の値にして、平均値、分散値を標準モデルの出力分布を1個の分布に表現したときの平均値、分散値にする。

のいずれかを利用して統計量の値を決定する。なお、利用する方法は、繰り返し回数R、HMM、HMMの状態ごとに異なっていてもよい。ここでは、第1の方法を用いる。

【0208】

統計量推定部306dは、このように推定した標準モデルの統計量を統計量記憶部306cに記憶する。そして、このような統計量の推定と統計量記憶部306cへの記憶をR(≥ 1)回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル322の統計量として出力する。

【0209】

次に、本実施の形態をPDAによる環境音識別に適用した具体例を説明する。

まず、参照モデル準備部302は、CD-ROMから環境音の識別に必要な参照モデルを読み出す。利用者は、識別を行う環境(利用情報)を考慮して、識別したい環境音を画面上から選択する。たとえば、「乗用車」を選択し、続いて、「警報音」、「赤ちゃんの声」、「電車の音」などを選択する。この選択に基づいて、参照モデル選択部305は、参照モデル記憶部303に記憶されている参照モデルの中から対応する参照モデルを選択する。そして、選択した参照モデル323を1つずつ用いて、標準モデル作成部306は、それぞれに対して標準モデルを作成する。

【0210】

続いて、利用者は、PDA301において、「らくらく情報提供」(環境音に基づく状況判断による情報提供)というアプリケーションプログラムを起動する。このアプリケーションは、環境音に基づいて状況判断を行い、利用者に適切な情報を提供するプログラムである。起動されると、PDA301の表示画面に「正確に判断」、「素早く判断」という表示がされる。これに対して、利用者はどちらかを選択する。

【0211】

そして、仕様情報作成部307は、その選択結果に基づいて、仕様情報を作成する。たとえば、「正確に判断」が選択された場合には、精度を高くするために、混合分布数を10個とする仕様情報を作成する。一方、「素早く判断」が選択された場合には、高速に処理するために、混合分布数を1個とする仕様情報を作成する。なお、複数のPDAが連携して処理できる場合などには、現在利用できるCPUパワーを判断し、そのCPUパワーに基づいて仕様情報を作成してもよい。

【0212】

このような仕様情報にしたがって、「乗用車」、「警報音」、「赤ちゃんの声」、「電車の音」などの1混合の標準モデルが作成される。そして、PDA301は、作成された標準モデルにより環境識別を行い、その識別結果に基づき、各種情報をPDAの画面に表示する。例えば、「乗用車」が近くにあると識別した場合は、道路地図を表示したり、「赤ちゃんの声」を識別した場合は、おもちゃ屋さんの広告を表示したりする。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを用いた、環境音識別に基づく情報提供が実現される。なお、アプリケーションの仕様に応じて標準モデルの複雑さを調節することができる。

【0213】

以上説明したように、本発明の第3の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率または尤度を最大化または極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

【0214】

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

なお、統計量推定部306dによる処理の繰り返し回数は、上記数25に示された尤度の大きさがある一定のしきい値以上になるまでの回数としてもよい。

【0215】

また、標準モデル322を構成するGMMは、雑音の種類ごとに異なる混合分布数の混合ガウス分布により構成してもよい。

また、識別モデルは、雑音モデルに限らず、話者を識別してもよいし、年齢などを識別してもよい。

【0216】

また、標準モデル322をCD-ROM、DVD-RAM、ハードディスクなどのストレージデバイスに記録してもよい。

また、参照モデル321をCD-ROMなどのストレージデバイスから読み込む代わりに、PDA301において雑音データから参照モデル321を作成してもよい。

【0217】

また、参照モデル準備部302は、必要に応じてCD-ROMなどのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部303に追加・更新してもよいし、参照モデル記憶部303に格納されている不要な参照モデルを削除してもよい。

また、参照モデル準備部302は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部303に追加・更新してもよい。

【0218】

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

【0219】

また、近傍指示パラメータGは、対象とする事象や標準モデルの出力分布によって異なってもよいし、繰り返し回数Rによって変化させてもよい。

【0220】

(第4の実施の形態)

図19は、本発明の第4の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ401に組み込まれた例が示されている。本実施の形態では顔認識用の標準モデルを作成する場合を例にして説明する。

【0221】

サーバ401は、通信システムにおけるコンピュータ装置等であり、事象の出力確率によって定義される顔認識用の標準モデルを作成する標準モデル作成装置として、カメラ411と、画像データ蓄積部412と、参照モデル準備部402と、参照モデル記憶部403と、利用情報受信部404と、参照モデル選択部405と、標準モデル作成部406と、書き込み部413とを備える。

【0222】

カメラ411により、顔の画像データが収集され、画像データ蓄積部412に顔画像データが蓄積される。参照モデル準備部402は、画像データ蓄積部412が蓄積した顔画像データを用いて話者ごとに参照モデル421を作成し、参照モデル記憶部403に記憶する。

【0223】

利用情報受信部404は、利用者が希望する顔認識の対象となる人間の年齢の年代と性別の情報を利用情報424として電話414により受信する。参照モデル選択部405は、利用情報受信部404が受信した利用情報424に基づいて、参照モデル記憶部403が記憶している参照モデル421の中から、利用情報424が示す年代と性別の話者に対応する参照モデル423を選択する。

【0224】

標準モデル作成部406は、参照モデル選択部405が選択した話者の顔画像の参照モデル423に対する確率または尤度を最大化または極大化するように標準モデル422を作成する処理部であり、第2の実施の形態における標準モデル作成部206と同一の機能を有するとともに、第1の実施の形態における第1近似部104eと第3の実施の形態における第2近似部306eの機能を有する。つまり、第1～第3の実施の形態で示された3種類の近似計算を組み合わせた計算を行う。

【0225】

書き込み部413は、標準モデル作成部406が作成した標準モデル422をCD-ROMなどのストレージデバイスに書き込む。

【0226】

次に、以上のように構成されたサーバ401の動作について説明する。

図20は、サーバ401の動作手順を示すフローチャートである。図21は、サーバ401の動作手順を説明するための参照モデルおよび標準モデルの一例を示す図である。

【0227】

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図20のステップS400）。つまり、カメラ411によりAさんからZさんの顔画像データを収集して画像データ蓄積部412に蓄積する。参照モデル準備部402は、画像データ蓄積部412が蓄積した顔画像データを用いて、話者ごとの参照モデル421をEMアルゴリズムにより作成する。ここでは参照モデル421はGMMで構成される。

【0228】

参照モデル記憶部403は、参照モデル準備部402が作成した参照モデル421を記憶する。ここでは、図21の参照モデル421に示されるように、AさんからZさんの全ての参照モデルが、混合分布数が5個のGMMにより構成される。特徴量として100次元（J=100）の画素の濃度値を用いる。

【0229】

次に、利用情報受信部404は、利用情報424である年代と性別の情報を電話414により受信する（図20のステップS401）。ここでは、利用情報424として、11歳から15歳の男性と22歳から26歳の女性である。参照モデル選択部405は、その利用情報424に基づいて、参照モデル記憶部403が記憶している参照モデル421から、利用情報424に対応する参照モデル423を選択する（図20のステップS402）。具体的には、図21の「選択された参照モデル423」に示されるように、ここでは、11歳から15歳の男性および22歳から26歳の女性の参照モデルを選択する。

【0230】

そして、標準モデル作成部406は、参照モデル選択部405が選択した話者の参照モデル423に対する確率または尤度を最大化または極大化するように標準モデル422を作成する（図20のステップS403）。ここでは、図21の標準モデル422に示されるように、2つの標準モデル422それぞれを、混合分布数が3個のGMMにより構成する。

【0231】

標準モデル422の作成方法は、基本的には、第2の実施の形態と同様に行われる。ただし、標準モデル422の統計量の推定における近似計算については、具体には、以下のようにして行われる。つまり、標準モデル作成部406は、内蔵の記憶部等を介することで、第1の実施の形態における第1近似部104eによる近似計算と同様の近似計算によって作成したモデルを初期値として、第2の実施の形態における一般近似部206eによ

る近似計算と同様の近似計算による計算を行い、その結果を初期値として第3の実施の形態における第2近似部306eによる近似計算と同様の近似計算を行う。

【0232】

書き込み部413は、標準モデル作成部406が作成した2つの標準モデル422をCD-ROMなどのストレージデバイスに書き込む（図20のステップS404）。

【0233】

利用者は、11歳から15歳の男性の標準モデルと22歳から26歳の女性の標準モデルが書き込まれたストレージデバイスを郵送で受け取る。

【0234】

次に、本実施の形態を、行動予測に基づいてお店などを紹介する情報提供システムに適用した具体例を説明する。この情報提供システムは、通信ネットワークで接続されたカーナビゲーション装置と情報提供サーバ装置から構成される。カーナビゲーション装置は、本実施の形態における標準モデル作成装置401によって予め作成された標準モデルを行動予測モデルとして利用することで、人の行動（つまり、車による行先等）を予測し、その行動に関連した情報（行先の近くに位置するレストランなどのお店の情報など）を提供する機能を備える。

【0235】

まず、利用者は、カーナビゲーション装置を用いて、電話回線414で接続されたサーバ401に対して、自分用の行動予測モデルの作成を依頼する。

【0236】

具体的には、利用者は、カーナビゲーション装置が表示する項目選択画面で、「らくらく推薦機能」のボタンを押す。すると、利用者の住所（利用場所）、年齢、性別、趣味などを入力する画面になる。

【0237】

ここでは、利用者はお父さんとお母さんとする。まず、お父さんの個人情報をカーナビゲーション装置の画面と対話しながら入力する。住所については、電話番号を入力することにより自動的に変換される。あるいは、カーナビゲーション装置において現在位置が表示されているときに「利用場所」のボタンを押すことで、その現在位置が利用場所として入力される。ここでは住所の情報を住所Aとする。年齢と性別については、「50代」、「男」を選択して入力する。趣味については、予め表示されたチェック項目があるので、利用者は、該当箇所をチェックする。ここではお父さんの趣味の情報を趣味情報Aとする。

【0238】

続いて、お母さんの個人情報についても同様に入力する。住所B、40代、女、趣味情報Bからなる個人情報が作成される。このような入力の結果は、図22の画面表示例に示されるとおりである。

【0239】

最後に、カーナビゲーション装置は、このようにして作成された個人情報を利用情報として、付属の電話回線414を用いて、情報提供サーバ装置であるサーバ401に転送する。

【0240】

次に、サーバ401は、転送されてきた個人情報（利用情報）に基づいて、お父さんとお母さんの2個の行動予測モデルを作成する。ここで、行動予測モデルは、確率モデルで表現され、その入力は、曜日、時刻、現在地などで、出力は、お店Aの情報を提示する確率、お店Bの情報を提示する確率、お店Cの情報を提示する確率、駐車場の情報を提示する確率などである。

【0241】

サーバ401の参照モデル記憶部403に記憶されている複数の参照モデルは、年代、性別、代表的な住所と趣味の傾向で作成した行動予測モデルである。サーバ401では、予め、カメラ411に代えて、カーナビゲーション装置の入力ボタン等を用いて各種個人

情報（上記入力及び出力についての情報）を入力することで、画像データ蓄積部412に各種個人情報を蓄積したうえで、参照モデル準備部402によって、画像データ蓄積部412に蓄積された個人情報から、複数種類の典型的な利用者ごとの参照モデル421が作成され、参照モデル記憶部403に格納されている。

【0242】

参照モデル選択部405は、個人情報（利用情報）を用いて、個人情報にふさわしい参照モデルを選択する。例えば、同じ町の、年代と性別が同じで、趣味のチェック項目が8割以上一致した参照モデルを選択する。サーバ401の標準モデル作成部406は、選択された参照モデルを統合した標準モデルを作成する。作成された標準モデルは書き込み部413により、メモリカードに記憶される。ここでは、お父さんとお母さんの2人の標準モデルが記憶される。メモリカードは、郵送で利用者に届けられる。

【0243】

利用者は、受け取ったメモリカードをカーナビゲーション装置に挿入し、画面に表示された「お父さん」と「お母さん」を選択することで、利用者を設定する。これによって、カーナビゲーション装置は、装着されたメモリカードに記憶された標準モデルを行動予測モデルとして使用することで、現在の曜日、時刻、場所などから、必要なタイミングでお店の情報などを提示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを行動予測モデルとして用いることで、人の行動（つまり、車による行先）を予測し、その行動に関連した情報を提供する情報提供システムが実現される。

【0244】

以上説明したように、本発明の第4の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率または尤度を最大化または極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい高精度な標準モデルが提供される。

【0245】

なお、標準モデル422を構成するGMMは、話者ごとに異なる分布数の混合ガウス分布により構成してもよい。

【0246】

また、参照モデル準備部402は、必要に応じて新たな参照モデルを作成して参照モデル記憶部403に追加・更新してもよいし、参照モデル記憶部403に格納されている不要な参照モデルを削除してもよい。

【0247】

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

【0248】

(第5の実施の形態)

図23は、本発明の第5の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ501に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（適応モデル）を作成する場合を例にして説明する。

【0249】

サーバ501は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象または事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部511と、音声データ蓄積部512と、参照モデル準備部502と、参照モデル記憶部503と、利用情報受信部504と、参照モデル選択部505と、標準モデル作成部506と、仕様情報受信部507と、書き込み部513とを備える。

【0250】

読み込み部511は、CD-ROMなどのストレージデバイスに書き込まれた子供、成人、高齢者の音声データを読み込み、音声データ蓄積部512に蓄積する。参照モデル準備部502は、音声データ蓄積部512が蓄積した音声データを用いて話者ごとに参照モデル521を作成する。参照モデル記憶部503は、参照モデル準備部502が作成した参照モデル521を記憶する。

【0251】

仕様情報受信部507は、仕様情報525を受信する。利用情報受信部504は、利用情報524である利用者の音声を受信する。参照モデル選択部505は、利用情報524である利用者の音声に音響的に近い話者の参照モデルを、参照モデル記憶部503が記憶している参照モデル521から選択する。

【0252】

標準モデル作成部506は、仕様情報525に基づいて、参照モデル選択部505が選択した話者の参照モデル523に対する確率または尤度を最大化または極大化するように標準モデル522を作成する処理部であり、第1の実施の形態における標準モデル作成部104と同一の機能を有する。書き込み部513は、標準モデル作成部506が作成した標準モデル522をCD-ROMなどのストレージデバイスに書き込む。

【0253】

次に、以上のように構成されたサーバ501の動作について説明する。

図24は、サーバ501の動作手順を示すフローチャートである。図25は、サーバ501の動作手順を説明するための参照モデルおよび標準モデルの一例を示す図である。

【0254】

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図24のステップS500）。つまり、読み込み部511は、CD-ROMなどのストレージデバイスに書き込まれた音声データを読み込み、音声データ蓄積部512に蓄積する。参照モデル準備部502は、音声データ蓄積部512が蓄積した音声データを用いて話者ごとに参照モデル521をバウム・ウェルチの再推定の方法により作成する。参照モデル記憶部503は、参照モデル準備部502が作成した参照モデル521を記憶する。

【0255】

参照モデル521は、音素ごとのHMMにより構成される。ここでは、図25の参照モデル521に示されるように、子供の各話者の参照モデルは、状態数3個、各状態は混合分布数が3個の混合ガウス分布によりHMMの出力分布が構成され、成人の各話者の参照モデルが、状態数3個、各状態は混合分布数が64個の混合ガウス分布によりHMMの出力分布が構成され、高齢者の各話者の参照モデルは、状態数3個、各状態は混合分布数が16個の混合ガウス分布によりHMMの出力分布が構成される。これは、子供の音声データが比較的少なく、成人の音声データが多いいためである。特徴量として25次元（J=25）のメルケプストラム係数が用いられる。

【0256】

次に、利用情報受信部504は、利用者の音声を、端末装置514から、利用情報524として受信する（図24のステップS501）。参照モデル選択部505は、利用情報524である利用者の音声に音響的に近い参照モデル523を、参照モデル記憶部503が記憶している参照モデル521から選択する（図24のステップS502）。具体的には、図25の「選択された参照モデル523」に示されるように、ここでは、近い話者10人（Ng=10）の参照モデルが選択される。

【0257】

そして、仕様情報受信部507は、利用者の要求に基づき仕様情報525を端末装置514から受信する（図24のステップS503）。ここでは、速い認識処理という仕様情報525を受信する。標準モデル作成部506は、仕様情報受信部507が受信した仕様情報525に基づいて、参照モデル選択部505が選択した話者の参照モデル523に対する確率または尤度を最大化または極大化するように標準モデル522を作成する（図24のステップS504）。具体的には、標準モデル522は、図25の標準モデル522

に示されるように、仕様情報525である速い認識処理という情報に基づいて、2混合（Mf=2）で、3状態のHMMより構成する。HMMは音素ごとに構成する。

【0258】

標準モデル522の作成方法は、第1の実施の形態と同様に行われる。

書き込み部513は、標準モデル作成部506が作成した標準モデル522をCD-ROMなどのストレージデバイスに書き込む（図24のステップS505）。

【0259】

次に、本実施の形態を、通信ネットワークを用いた音声認識によるゲームに適用した具体例を説明する。ここでは、サーバ501は、作成した標準モデルを用いて音声認識を行う音声認識部を備えるものとする。また、端末装置514として、PDAとする。これらは、通信ネットワークで接続されている。

【0260】

サーバ501では、読み込み部511、音声データ蓄積部512及び参照モデル準備部502により、音声データをCDやDVDなどで入手したタイミングで参照モデルを逐次準備している。

【0261】

利用者は、PDA（端末装置514）において、音声認識を利用したゲームプログラム、ここでは、「アクションゲーム」を立ち上げる。すると、「『アクション』と発声してください」と表示されるので、利用者は、「アクション」と発声する。その音声は、利用情報として、PDA（端末装置514）からサーバ501に送信され、サーバ501の利用情報受信部504及び参照モデル選択部505により、参照モデル記憶部503に記憶された複数の参照モデルの中から利用者に合った参照モデルを選択する。

【0262】

また、利用者は、速くリアクションしてほしいので、PDA（端末装置514）の設定画面において「高速に認識する」と設定する。その設定内容は、仕様情報として、PDA（端末装置514）からサーバ501に送信され、サーバ501においては、このような仕様情報及び選択された参照モデルに基づいて、標準モデル作成部506により、2混合の標準モデルが作成される。

【0263】

利用者は、アクションゲームにおいて、PDAのマイクに「右に移動」、「左に移動」などのコマンドを発声する。入力された音声は、サーバへ送信され、既に作成された標準モデルを利用した音声認識が行われる。その認識結果は、サーバ501からPDA（端末装置514）に送信され、PDA（端末装置514）において、送信されてきた認識結果に基づいて、アクションゲームのキャラクタが動く。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声によるアクションゲームが実現される。

【0264】

また、同様にして、本実施の形態を別のアプリケーション、例えば、通信ネットワークを用いた翻訳システムに適用することもできる。たとえば、利用者は、PDA（端末装置514）において、「音声翻訳」というアプリケーションプログラムを立ち上げる。すると、「『翻訳』と発声してください」と表示される。利用者は、「翻訳」と発声する。その音声は、利用情報として、PDA（端末装置514）からサーバ501に送信される。また、利用者は、正確に認識してほしいので、そのアプリケーションにおいて、「正確に認識してほしい」旨を指示する。その指示は、仕様情報として、PDA（端末装置514）からサーバ501に送信される。サーバ501では、送信されてきた利用情報及び仕様情報に従って、たとえば、100混合の標準モデルが作成される。

【0265】

利用者は、PDA（端末装置514）のマイクに向かって「おはようございます」と発声する。入力された音声はPDA（端末装置514）からサーバ501に送信され、サーバ501で「おはようございます」と認識された後に、その認識結果がPDA（端末装置

514)に返信される。PDA(端末装置514)は、サーバ501から受信した認識結果を英語に翻訳し、その結果「GOOD MORNING」を画面に表示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声による翻訳装置が実現される。

【0266】

以上説明したように、本発明の第5の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率または尤度を最大化または極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

【0267】

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

【0268】

また、参照モデル準備部502において、参照モデルごとにデータ数に適した混合分布数の精度の高い参照モデルを準備でき、精度の高い参照モデルを用いて標準モデルを作成できる。このため精度の高い標準モデルの利用が可能となる。

【0269】

なお、標準モデル522は、音素ごとにHMMを構成するに限らず、文脈依存のHMMで構成してもよい。

【0270】

また、標準モデル522を構成するHMMは、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

また、標準モデル522を用いて、サーバ501において音声認識を行ってもよい。

【0271】

また、参照モデル準備部502は、必要に応じて新たな参照モデルを作成して参照モデル記憶部503に追加・更新してもよいし、参照モデル記憶部503に格納されている不要な参照モデルを削除してもよい。

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

【0272】

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

【0273】

(第6の実施の形態)

図26は、本発明の第6の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ601に組み込まれた例が示されている。本実施の形態では意図理解のための標準モデル(嗜好モデル)を作成する場合を例にして説明する。

【0274】

サーバ601は、通信システムにおけるコンピュータ装置等であり、事象の出力確率によって定義される意図理解用の標準モデルを作成する標準モデル作成装置として、読み込み部611と、参照モデル準備部602と、参照モデル記憶部603と、利用情報受信部604と、参照モデル選択部605と、標準モデル作成部606と、仕様情報作成部607とを備える。

【0275】

読み込み部611は、CD-ROMなどのストレージデバイスに書き込まれた年齢別の話者Aさんから話者Zさんの嗜好モデルを読み込み、参照モデル準備部602は、読み込まれた参照モデル621を参照モデル記憶部603へ送信し、参照モデル記憶部603は、参照モデル621を記憶する。

【0276】

仕様情報作成部607は、普及しているコンピュータのCPUパワーに合わせて仕様情

報625を作成する。利用情報受信部604は、端末装置614から利用情報624を受信する。参照モデル選択部605は、利用情報受信部604が受信した利用情報624に基づいて、参照モデル記憶部603が記憶している参照モデル621の中から、利用情報624に対応した参照モデル623を選択する。

【0277】

標準モデル作成部606は、仕様情報作成部607が作成した仕様情報625に基づいて、参照モデル選択部605が選択した参照モデル623に対する確率または尤度を最大化または極大化するように標準モデル622を作成する処理部であり、第2の実施の形態における標準モデル作成部206と同一の機能を有するとともに、第3の実施の形態における第2近似部306eの機能を有する。つまり、第2および第3の実施の形態で示された2種類の近似計算を組み合わせた計算を行う。

【0278】

次に、以上のように構成されたサーバ601の動作について説明する。

図27は、サーバ601の動作手順を示すフローチャートである。図28は、サーバ601の動作手順を説明するための参照モデルおよび標準モデルの一例を示す図である。

【0279】

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図27のステップS600）。つまり、読み込み部611は、CD-ROMなどのストレージデバイスに書き込まれた年齢別の話者Aさんから話者Zさんの嗜好モデルを読み込み、参照モデル準備部602は、読み込まれた参照モデル621を参照モデル記憶部603へ送信し、参照モデル記憶部603は、参照モデル621を記憶する。

【0280】

参照モデル621は、GMMより構成される。ここでは、図28の参照モデル621に示されるように、混合分布数が3個のGMMにより構成される。学習データとして、趣味、性格などを数値化した5次元（J=5）の特徴量を用いる。参照モデルの準備は、標準モデルの作成が要求される以前に行う。

【0281】

次に、利用情報受信部604は、嗜好モデルを作成したい年齢層である利用情報624を受信する（図27のステップS601）。ここでは、20代、30代、40代の年代別の嗜好モデルを利用するという利用情報624である。参照モデル選択部605は、図28の「選択された参照モデル623」に示されるように、利用情報受信部604が受信した利用情報624が示す年代の話者の嗜好モデルを、参照モデル記憶部603が記憶している参照モデル621から選択する（図27のステップS602）。

【0282】

そして、仕様情報作成部607は、普及しているコンピュータのCPUパワー、記憶容量などに基づき仕様情報625を作成する（図27のステップS603）。ここでは、通常速度の認識処理という仕様情報625を作成する。

【0283】

標準モデル作成部606は、仕様情報作成部607が作成した仕様情報625に基づいて、参照モデル選択部605が選択した話者の参照モデル623に対する確率または尤度を最大化または極大化するように標準モデル622を作成する（図27のステップS604）。ここでは、標準モデル622は、図28の標準モデル622に示されるように、仕様情報625である通常速度の認識処理という情報に基づいて3混合（Mf=3）のGMより構成する。

【0284】

標準モデル622の作成方法は、基本的には、第2の実施の形態と同様に行われる。ただし、標準モデル622の統計量の推定における近似計算については、具体には、以下のようにして行われる。つまり、標準モデル作成部606は、内蔵の記憶部等を介することで、第2の実施の形態における一般近似部206eによる近似計算と同様の近似計算による計算を行い、その結果を初期値として第3の実施の形態における第2近似部306e

による近似計算と同様の近似計算を行う。

【0285】

次に、本実施の形態を情報検索装置に適用した具体例を説明する。ここでは、参照モデルは、入力が検索キーワードであり、出力が検索ルールA、検索ルールBなどを利用する確率である。異なる検索ルールを用いると、表示される検索結果が異なってくる。また、サーバ601の参照モデル記憶部603に準備される参照モデルは、代表的な特徴をもつ話者のモデルとする。

【0286】

まず、利用者は、サーバ601に付属しているリモコン（端末装置614）を用いて利用情報を入力する。利用情報は、年齢、性格、性別、趣味などである。また、「子供」、「俳優」、「高校生」などの所定のグループを識別する情報であってもよい。

【0287】

続いて、利用者は、選択画面で、「カーナビゲーション装置用」、「携帯電話用」、「パソコン用」、「テレビ用」などから1つの利用機器を選択する。サーバ601の仕様情報作成部607は、利用機器のC.P.Uパワー、記憶容量に基づいて仕様情報を作成する。ここでは、「テレビ用」が選択されたとし、C.P.Uパワーと記憶容量が小さい旨の仕様情報625が作成され、その仕様情報625に基づいて、標準モデル作成部606によって、小さいC.P.Uパワーでも動作する3混合の標準モデルが作成される。作成された標準モデルはメモリカードに書き込まれ、そのメモリカードは利用者によってテレビに挿入される。

【0288】

利用者は、テレビに表示されたE.P.Gなどで、おすすめ番組を検索するために検索キーワードを入力する。すると、テレビは、メモリカードに記録された標準モデルを用いて、検索キーワードに合った検索ルールを決定し、その検索ルールに沿って番組を検索し、利用者の嗜好にあった番組として表示する。このようにして、本実施の形態における標準モデル作成装置によって作成された標準モデルを用いた便利な検索装置が実現される。

【0289】

以上説明したように、本発明の第6の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率または尤度を最大化または極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

【0290】

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

【0291】

なお、標準モデル622を構成するGMMは、話者ごとに異なる分布数の混合ガウス分布により構成してもよい。

【0292】

また、参照モデル準備部602は、必要に応じてCD-ROMなどのストレージデバイスから読み込まれた新たな参照モデルを参照モデル記憶部603に追加・更新してもよいし、参照モデル記憶部603に格納されている不要な参照モデルを削除してもよい。

また、参照モデルおよび標準モデルのGMMはペイジアンネットの一部を表現するものでもよい。

【0293】

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

【0294】

(第7の実施の形態)

図29は、本発明の第7の実施の形態における標準モデル作成装置の全体構成を示すブ

ロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ701に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（適応モデル）を作成する場合を例にして説明する。

【0295】

サーバ701は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象または事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部711と、参照モデル準備部702と、参照モデル記憶部703と、利用情報受信部704と、参照モデル選択部705と、標準モデル作成部706と、仕様情報受信部707と、標準モデル記憶部708と、標準モデル送信部709とを備える。

【0296】

参照モデル準備部702は、読み込み部711が読み込んだ、CD-ROMなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部703へ送信し、参照モデル記憶部703は、送信された参照モデル721を記憶する。

【0297】

仕様情報受信部707は、端末装置712から仕様情報725を受信する。利用情報受信部704は、端末装置712から利用情報724である雑音下で発声した利用者の音声を受信する。参照モデル選択部705は、利用情報724である利用者の音声に音響的に近い話者・雑音・声調子の参照モデル723を、参照モデル記憶部703が記憶している参照モデル721の中から選択する。

【0298】

標準モデル作成部706は、仕様情報受信部707が受信した仕様情報725に基づいて、参照モデル選択部705が選択した参照モデル723に対する確率または尤度を最大化または極大化するように標準モデル722を作成する処理部であり、第2の実施の形態における標準モデル作成部206と同一の機能を有する。標準モデル記憶部708は、仕様情報725に基づいた1もしくは複数の標準モデルを記憶する。標準モデル送信部709は、利用者の端末装置712から仕様情報と標準モデルの要求信号を受信すると、その仕様情報に適した標準モデルを端末装置712へ送信する。

【0299】

次に、以上のように構成されたサーバ701の動作について説明する。

図30は、サーバ701の動作手順を示すフローチャートである。図31は、サーバ701の動作手順を説明するための参照モデルおよび標準モデルの一例を示す図である。

【0300】

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図30のステップS700）。つまり、参照モデル準備部702は、読み込み部711が読み込んだ、CD-ROMなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部703へ送信し、参照モデル記憶部703は、送信された参照モデル721を記憶する。ここでは、参照モデル721は、話者・雑音・声の調子ごとに、音素ごとのHMMにより構成される。また、各参照モデルは、図31の参照モデル721に示されるように、状態数3個、各状態は混合分布数が128個の混合ガウス分布によりHMMの出力分布が構成される。特徴量として25次元（J=25）のマルケプストラム係数が用いられる。

【0301】

次に、利用情報受信部704は、利用者Aの雑音下での音声を端末装置712から利用情報724として受信する（図30のステップS701）。参照モデル選択部705は、利用情報724である利用者Aの音声に音響的に近い参照モデル723を、参照モデル記憶部703が記憶している参照モデル721の中から選択する（図30のステップS702）。具体的には、図31の「選択された参照モデル723」に示されるように、ここでは、近い話者100人（Ng=100）の参照モデルが選択される。

【0302】

そして、仕様情報受信部707は、利用者Aの要求に基づき仕様情報725を端末装置712から受信する（図30のステップS703）。ここでは、高い認識精度という仕様情報725を受信する。標準モデル作成部706は、仕様情報725に基づいて、参照モデル選択部705が選択した参照モデル723に対する確率または尤度を最大化または極大化するように標準モデル722を作成する（図30のステップS704）。具体的には、標準モデル722は、図31の標準モデル722に示されるように、仕様情報725である高い認識精度という情報に基づいて、64混合（Mf=64）で、3状態のHMMにより構成する。HMMは音素ごとに構成する。

【0303】

標準モデル722の作成方法は、第2の実施の形態と同様に行われる。

標準モデル記憶部708は、仕様情報725に基づいた1もしくは複数の標準モデル722を記憶する。ここでは、以前に作成した標準モデルである利用者Bの16混合のHMMがすでに記憶されており、新たに利用者Aの64混合のHMMが記憶される。

【0304】

利用者Aは、端末装置712からサーバ701の標準モデル送信部709へ、仕様情報である利用者Aと雑音の種類と標準モデルの要求信号とを送信する（図30のステップS706）。標準モデル送信部709は、利用者Aが送信した仕様情報と標準モデルの要求信号とを受信すると、その端末装置712へ、仕様に適した標準モデルを端末装置712へ送信する（図30のステップS707）。ここでは、先ほど作成した利用者Aの標準モデル722を端末装置712へ送信する。

【0305】

利用者Aは端末装置712において受信した標準モデル722を用いて音声認識を行う（図30のステップS708）。

【0306】

次に、本実施の形態を、通信ネットワークで接続されたカーナビゲーション装置（端末装置712）とサーバ装置（サーバ701；標準モデル作成装置）から構成される音声認識システムに適用した具体例を説明する。

【0307】

まず、利用者は、カーナビゲーション装置（端末装置712）の画面にて「自分の音声モデルを獲得」する旨のボタンを選択する。すると、「名前を入力」と表示されるので、ボタン操作により自分の名前を入力する。次に、「『音声』と発声してください」と表示されるので、利用者は、カーナビゲーション装置付属のマイクに向かって「音声」と発声する。これらの情報（利用者の名前、雑音下での音声）は、利用情報として、カーナビゲーション装置（端末装置712）からサーバ701に送信される。

【0308】

同様にして、利用者は、カーナビゲーション装置（端末装置712）の画面にて「高精度の音声認識」のボタンを選択する。すると、その選択情報は、仕様情報として、カーナビゲーション装置（端末装置712）からサーバ701に送信される。

【0309】

サーバ701は、それらの利用情報及び仕様情報に基づいて、利用者にふさわしい音声認識用の標準モデルを作成し、作成した標準モデルを利用者の名前と対応づけて標準モデル記憶部708に格納しておく。

【0310】

次回にカーナビゲーション装置（端末装置712）を起動すると、「名前を入力」と表示されるので、利用者は、名前を入力する。すると、その名前がサーバ701に送信され、標準モデル722に格納された対応する標準モデルが標準モデル送信部709によってサーバ701から端末装置712に送信される。名前（利用者）に対応した標準モデルをサーバ701からダウンロードした端末装置712は、その標準モデルを用いて、利用者に対する音声認識を行い、音声による目的地設定などを行う。このようにして、本実施の

形態における標準モデル作成装置によって作成された標準モデルを音声認識に用いることで、音声によってカーナビゲーション装置を操作することが可能となる。

【0311】

以上説明したように、本発明の第7の実施の形態によれば、利用情報に基づいて選択された複数の参照モデルに対する確率または尤度を最大化または極大化するように標準モデルの統計量を計算して標準モデルが作成されるので、利用状況によりふさわしい精度の高い標準モデルが提供される。

【0312】

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

また、標準モデル記憶部708は、複数の標準モデルを記憶することができるため、必要に応じてすぐに標準モデルが提供される。

【0313】

また、標準モデル送信部709により、標準モデルが端末装置712へ送信されるので、端末装置712とサーバ701が空間的に離れた場所に設置してある場合に、端末装置712は、容易にサーバ701が作成した標準モデルを利用することできる。

【0314】

なお、標準モデル722は、音素ごとにHMMを構成するに限らず、文脈依存のHMMで構成してもよい。

【0315】

また、標準モデル722を構成するHMMは、状態ごとに異なる混合数の混合ガウス分布により構成してもよい。

また、標準モデル722を用いて、サーバ701において音声認識を行い、認識結果を端末装置712へ送信してもよい。

【0316】

また、参照モデル準備部702は、必要に応じて新たな参照モデルを作成して参照モデル記憶部703に追加・更新してもよいし、参照モデル記憶部703に格納されている不要な参照モデルを削除してもよい。

【0317】

また、参照モデル準備部702は、必要に応じて、通信路を介して新たな参照モデルを参照モデル記憶部703に追加・更新してもよい。

【0318】

また、標準モデルを作成したのちに、さらにデータにより学習してもよい。
また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

【0319】

(第8の実施の形態)

図32は、本発明の第8の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置が携帯電話901に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデルを作成する場合を例にして説明する。

【0320】

携帯電話901は、携帯情報端末であり、事象の集合と事象または事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、参照モデル受信部909と、参照モデル準備部902と、参照モデル記憶部903と、利用情報作成部904と、参照モデル選択部905と、類似度情報作成部908と、標準モデル作成部906と、仕様情報作成部907と、マイク912と、音声認識部913とを備える。

【0321】

利用情報作成部904は、利用情報924を携帯電話901の画面とキーを利用して作

成する。

【0322】

仕様情報作成部907は、携帯電話901の仕様に基づき仕様情報925を作成する。ここで、仕様情報とは、作成する標準モデルの仕様に関する情報であり、ここでは、携帯電話901が備えるC P Uの処理能力に関する情報である。

【0323】

類似度情報作成部908は、利用情報924と仕様情報925と参照モデル記憶部903が記憶した参照モデル921に基づいて、類似度情報926を作成して参照モデル準備部に送信する。

【0324】

参照モデル準備部902は、類似度情報926に基づいて、参照モデルを準備するか否かを決定する。参照モデル準備部902は、参照モデルを準備すると決定した場合に、利用情報924と仕様情報925を参照モデル受信部909に送信する。

【0325】

参照モデル受信部909は、利用情報924と仕様情報925に対応した参照モデルを、サーバ装置910から受信して参照モデル準備部902に送信する。

【0326】

参照モデル準備部902は、参照モデル受信部909が送信した参照モデルを参照モデル記憶部903に記憶する。

【0327】

参照モデル選択部905は、利用情報924に対応した参照モデル923を、参照モデル記憶部903が記憶している参照モデル921の中から選択する。

【0328】

標準モデル作成部906は、仕様情報作成部907で作成された仕様情報925に基づいて、参照モデル選択部905が選択した参照モデル923に対する確率または尤度を最大化または極大化するように標準モデル922を作成する処理部であり、標準モデルの構造（ガウス分布の混合分布数など）を決定する標準モデル構造決定部906aと、標準モデルを計算するための統計量の初期値を決定することで初期標準モデルを作成する初期標準モデル作成部906bと、決定された初期標準モデルを記憶する統計量記憶部906cと、統計量記憶部906cに記憶された初期標準モデルに対して、第3近似部906eによる近似計算等を用いることにより、参照モデル選択部905が選択した参照モデル923に対する確率または尤度を最大化または極大化するような統計量を算出する（最終的な標準モデルを生成する）統計量推定部906dとからなる。

【0329】

音声認識部913は、標準モデル作成部906で作成された標準モデル922を用いて、マイク912から入力された利用者の音声を認識する。

【0330】

次に、以上のように構成された携帯電話901の動作について説明する。

図33は、携帯電話901の動作手順を示すフローチャートである。

いま、参照モデル記憶部903には、あらかじめ参照モデル921として子供用モデルが記憶されているとする。その参照モデル921は、音素ごとのHMMにより構成される。参照モデル921の一例を図34に示す。ここでは、子供用参照モデルのイメージ図が示されている。これらの参照モデルは、状態数3個、各状態は分布数が16個の混合ガウス分布によりHMMの出力分布が構成される。特徴量として、12次元のメルケプストラム係数、12次元のデルタメルケプストラム係数、デルタパワーの合計25次元（J=25）の特徴量が用いられる。

【0331】

まず、利用情報作成部904は、利用者の属するカテゴリである利用情報924を作成する（ステップS900）。図36は、利用情報924の作成例を示す図である。図36(a)に携帯電話901の選択画面の一例を示す。ここでは、「4：成人」のボタンを押

することにより、この携帯電話901が成人女性と成人男性に利用されることが選択されている。別の一例を図36(b)に示す。ここでは、「メニュー」ボタンを押しながら音声を入力している。その利用者の音声は、特徴量に変換されることで、利用情報924である”利用者の音声データ”が作成される。

【0332】

一方、仕様情報作成部907は、携帯電話901の仕様に基づき、仕様情報925を作成する(ステップS901)。ここでは、携帯電話901のメモリ容量の大きさに基づいて「混合分布数16」という仕様情報925を作成する。

【0333】

次に、類似度情報作成部908は、利用情報924と仕様情報925と参照モデル記憶部903が記憶した参照モデル921に基づいて、類似度情報926を作成して(ステップS902)、類似度情報926を参照モデル準備部902に送信する。ここでは、参照モデル記憶部903に存在する参照モデル921は、混合分布数3の子供用モデル(図34を参照)のみであり、利用情報924である「成人」(図36(a)に対応)と仕様情報925である「混合分布数16」に対応する参照モデルが参照モデル記憶部903に存在しないため、「類似した参照モデルが存在しない」という類似度情報926を作成して、類似度情報926を参照モデル準備部902に送信する。別の一例では、利用情報924は”利用者の音声データ”(図36(b)に対応)であり、利用者の音声データを参照モデル記憶部903が記憶している子供用モデルに入力して類似度情報926を作成する。ここでは、子供用モデルに対する尤度が所定のしきい値以下であるため、「類似した参照モデルが存在しない」という類似度情報926を作成して参照モデル準備部902に送信する。

【0334】

続いて、参照モデル準備部902は、類似度情報926に基づいて、参照モデルを準備するか否かを決定する(ステップS903)。ここでは、「類似した参照モデルが存在しない」ため、図37(a)の携帯電話901の画面表示例に示すように利用者に参照モデルの準備を促す。ここで、利用者が「メモ」ボタンを押して参照モデルの準備を要求した場合に、参照モデル準備部902は、参照モデルを準備すると決定して、利用情報924と仕様情報925を参照モデル受信部909に送信する。別の一例では、「類似した参照モデルが存在しない」ため、参照モデル準備部902は、自動的に参照モデルを準備すると決定して、利用情報924と仕様情報925を参照モデル受信部909に送信する。この場合の携帯電話901の画面の一例を図37(b)に示す。

【0335】

これに対して、参照モデル受信部909は、利用情報924と仕様情報925に対応した参照モデルをサーバ装置910から受信して参照モデル準備部902に送信する。ここでは、参照モデル受信部909は、利用情報924である「成人」(図36(a)に対応)と仕様情報925である「混合分布数16」に対応する参照モデルである、”混合分布数16の成人女性用モデル”と”混合分布数16の成人男性用モデル”的2個の参照モデルをサーバ装置910から受信する。

【0336】

そして、参照モデル準備部902は、参照モデル受信部909が送信した参照モデルを参照モデル記憶部903に記憶することによって参照モデルを準備する(ステップS904)。図35にその参照モデルの一例を示す。ここでは、成人男性用、成人女性用、子供用の参照モデルのイメージ図が示されている。

【0337】

次に、参照モデル選択部905は、利用情報924である「成人」に対応した同じカテゴリに属する”混合分布数16の成人女性用モデル”と”混合分布数16の成人男性用モデル”的2個の参照モデルを参照モデル記憶部903が記憶している参照モデル921の中から選択する(ステップS905)。別の一例では、参照モデル選択部905は、利用情報924である”利用者の音声データ”と音響的に近い(尤度が大きい)”混合分布数16

の成人女性用モデル”と”混合分布数16の成人男性用モデル”的2個の参照モデルを参照モデル記憶部903が記憶している参照モデル921の中から選択する。

【0338】

続いて、標準モデル作成部906は、作成された仕様情報925に基づいて、参照モデル選択部905が選択した参照モデル923に対する確率または尤度を最大化または極大化するように標準モデル922を作成する（ステップS906）。

【0339】

最後に、音声認識部913は、標準モデル作成部906によって作成された標準モデル922に従って、マイク912から入力された利用者の音声を認識する（ステップS907）。

【0340】

次に、図33におけるステップS906（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図4に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

【0341】

まず、標準モデル構造決定部906aは、標準モデルの構造を決定する（図4のステップS102）。ここでは、標準モデルの構造として、仕様情報925である「混合分布数16」に基づいて、音素ごとのHMMにより構成し、状態数を3とし、各状態における出力分布の混合分布数を16個（Mf=16）と決定する。

【0342】

次に、初期標準モデル作成部906bは、標準モデルを計算するための統計量の初期値を決定する（図4のステップS102b）。ここでは、選択された参照モデル923である”混合分布数16の成人女性用モデル”を統計量の初期値として統計量記憶部906cに記憶する。別の一例では、選択された参照モデル923である”混合分布数16の成人男性用モデル”を統計量の初期値として統計量記憶部906cに記憶する。具体的には、初期標準モデル作成部906bは、上記数13に示される出力分布を生成する。

【0343】

そして、統計量推定部906dは、参照モデル選択部905が選択した2個の参照モデル923を用いて、統計量記憶部906cに記憶された標準モデルの統計量を推定する（図4のステップS102c）。つまり、2個（Ng=2）の参照モデル923における出力分布、即ち、上記数19に示される出力分布に対する標準モデルの確率（ここでは、上記数25に示される尤度logP）を極大化もしくは最大化するような標準モデルの統計量（上記数16に示される混合重み係数、上記数17に示される平均値、および、上記数18に示される分散値）を推定する。ただし、本実施の形態では、上記数19に示された出力分布における数21は、16（各参照モデルの混合分布数）である。

【0344】

具体的には、上記数26、数27および数28に示される式に従って、それぞれ、標準モデルの混合重み係数、平均値および分散値を算出する。

【0345】

このとき、統計量推定部906dの第3近似部906eは、標準モデルの各ガウス分布はお互いに影響を与えないと仮定して、数53の近似式を用いる。また、繰り返し回数Rが1回目の場合には、数54に示される標準モデルのガウス分布の近傍の数55とは、数54が示す出力分布とのマハラノビス距離、カルバック・ライプラー（KL）距離などの分布間距離が最も近いものと2番目に近いものの2個（近傍指示パラメータG=2）の数56に示される参照モデル923のガウス分布が存在する空間であると近似する。一方、繰り返し回数Rが2回目以上の場合には、数54に示される標準モデルのガウス分布の近傍の数55とは、数54が示す出力分布とのマハラノビス距離、カルバック・ライプラー（KL）距離などの分布間距離が最も近いもの1個（近傍指示パラメータG=1）の数56に示される参照モデル923のガウス分布が存在する空間であると近似する。

【0346】

以上の第3近似部906eによる近似式を考慮してまとめると、統計量推定部906dでの計算式は、次の通りになる。つまり、統計量推定部906dは、数59、数60および数61に示される式に従って、それぞれ、混合重み係数、平均値および分散値を算出し、それらのパラメータによって特定される標準モデルを最終的な標準モデル922として生成する。ただし、第3の実施の形態における第2の方法である、混合重み係数の値をゼロにして、平均値をゼロ、分散値を1にする方法を用いる。また、繰り返し回数に対応して近傍指示パラメータGの値は異なる。なお、近傍指示パラメータGの値に依存して、上記の方法を、第3の実施の形態における第1から第3の方法のいずれかに決定してもよい。

【0347】

統計量推定部906dは、このように推定した標準モデルの統計量を統計量記憶部906cに記憶する。そして、このような統計量の推定と統計量記憶部906cへの記憶をR(≥ 1)回、繰り返す。その結果得られた統計量を最終的に生成する標準モデル922の統計量として出力する。

【0348】

図38に、第3近似部906eを用いて作成した標準モデル922を用いた認識実験の結果を示す。縦軸に成人(男性と女性)の認識率(%)、横軸に繰り返し回数Rを示す。繰り返し回数R=0とは、学習を行う前の初期標準モデル作成部906bが作成した初期モデルにより認識した結果である。また、繰り返し回数R=1のときは、近傍指示パラメータG=2とし、繰り返し回数R=2~5のときは、近傍指示パラメータG=1とした。

【0349】

グラフ「データ」は、数日間かけて音声データより学習した場合の結果を表しており、グラフ「女性」、グラフ「男性」は、それぞれ、初期モデルを成人女性、成人男性としたときの結果を表している。参照モデルによる本発明による学習時間は数十秒のオーダーであった。実験結果より、短時間に高い精度の標準モデルが作成できていることがわかる。

【0350】

ここで、参考のために、図39に、第3の実施の形態における第2近似部306eにより作成された標準モデルによる認識率を示す。本実施の形態における第3近似部906eと異なるのは、繰り返し回数Rによらず近傍指示パラメータG=1であるということである。実験結果より、初期モデルとして成人女性を選択すると良好な結果が得られることがわかる。また、初期モデルとして成人男性を選択すると、精度が少し劣化していることがわかる。図38の結果とあわせると、第3近似部906eによる標準モデルは初期モデルに依存せずに高い精度の標準モデルが作成できていることがわかる。

【0351】

以上説明したように、本発明の第8の実施の形態によれば、類似度情報に基づいて参照モデルを準備するため、利用情報および仕様情報にふさわしい参照モデルを必要なタイミングで準備することができる。また、近傍指示パラメータGを繰り返し回数Rによって変化させることで、初期モデルにかかわらず精度の高い標準モデルを提供することができる。

【0352】

なお、統計量推定部906dによる処理の繰り返し回数は、上記数25に示された尤度の大きさがある一定のしきい値以上になるまでの回数としてもよい。

また、標準モデル922は、音素ごとにHMMを構成するに限らず、文脈依存のHMMで構成してもよい。

【0353】

また、標準モデル作成部906は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

【0354】

また、標準モデル922を構成するHMMは、音素ごとに異なる状態数により構成して

もよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

【0355】

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

【0356】

(第9の実施の形態)

図40は、本発明の第9の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がPDA (Personal Digital Assistant) 1001に組み込まれた例が示されている。以下、本実施の形態では音声認識用の標準モデルを作成する場合を例にして説明する。

【0357】

PDA 1001は、携帯情報端末であり、事象の集合と事象または事象間の遷移の出力確率で表現された隠れマルコフモデルによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、参照モデル記憶部1003と、標準モデル作成部1006と、アプリ・仕様情報対応データベース1014と、マイク1012と、音声認識部1013とを備える。標準モデル作成部1006は、標準モデル構造決定部1006aと、初期標準モデル作成部1006bと、統計量記憶部306cと、統計量推定部306dとを備える。

【0358】

標準モデル作成部1006は、送信されたアプリ起動情報1027（ここでは、起動したアプリケーションのID番号）に基づいて、アプリ・仕様情報対応データベース1014を用いて、仕様情報1025を取得する。図41は、仕様情報対応データベース1014のデータ例を示す。仕様情報対応データベース1014には、アプリケーション（ID番号および名前）に対応する仕様情報（ここでは、混合分布数）が登録されている。

【0359】

標準モデル作成部1006は、取得した仕様情報1025に基づいて、参照モデル記憶部1003が記憶した1個の参照モデル1023に対する確率または尤度を最大化または極大化するように標準モデル1022を作成する処理部であり、第3の実施の形態における第2近似部306eの機能を有する。

【0360】

音声認識部1013は、標準モデル作成部1006で作成された標準モデル1022を用いて、マイク1012から入力された利用者の音声を認識する。

【0361】

次に、以上のように構成されたPDA 1001の動作について説明する。

図42は、PDA 1001の動作手順を示すフローチャートである。

【0362】

ここで、参照モデル記憶部1003には、あらかじめ多くの混合分布数をもつ利用者用モデルが参照モデル1021として1個、記憶されているとする。参照モデル1021は、音素ごとのHMMにより構成される。参照モデル1021の一例を図43に示す。この参照モデルは、状態数3個、各状態は分布数が300個の混合ガウス分布によりHMMの出力分布が構成される。特徴量として、12次元のメルケプストラム係数、12次元のデルタメルケプストラム係数、デルタパワーの合計25次元（J=25）の特徴量が用いられる。

【0363】

まず、利用者は、例えば「株取引」というアプリケーションを起動する（ステップS1000）。

【0364】

これに対して、標準モデル作成部1006は、アプリ起動情報として起動されたアプリケーションのID「3」を受信する（ステップS1001）。そして、アプリ・仕様情報

対応データベース1014を用いてID「3」に対応する仕様情報1025である「混合分布数126」に基づいて、標準モデル1022を作成する（ステップS1002）。具体的には、標準モデル1022として、混合分布数126（Mf=126）で、3状態の文脈依存型のHMMにより構成する。

【0365】

次に、標準モデル作成部1006は、仕様情報1025を受信して（ステップS1001）、仕様情報1025に基づいて標準モデルを作成する（ステップS1002）。

【0366】

最後に、音声認識部1013は、標準モデル作成部1006によって作成された標準モデル1022に従って、マイク1012から入力された利用者の音声を認識する（ステップS1003）。

【0367】

次に、図42におけるステップS1002（標準モデルの作成）の詳細な手順を説明する。手順の流れは、図4に示されたフローチャートと同様である。ただし、採用する標準モデルの構造や具体的な近似計算等が異なる。

【0368】

まず、標準モデル構造決定部1006aは、アプリ起動情報1027としてアプリケーションID「3」を受信した後に、アプリ・仕様情報対応データベース1014を用いてID「3」に対応した仕様情報1025（「混合分布数126」）を参照することにより、標準モデルの構造を混合分布数126（Mf=126）で、3状態の文脈依存型のHMMと決定する（図4のステップS102a）。

【0369】

そして、初期標準モデル作成部1006bは、標準モデル構造決定部1006aが決定した標準モデルの構造に基づいて、標準モデルを計算するための統計量の初期値を決定する（図4のステップS102b）ここでは、k-means法とマハラノビス汎距離を用いた方法により、後述するクラスタリングを行ったものを統計量の初期値として統計量記憶部306cに記憶する。

【0370】

そして、統計量推定部306dは、参照モデル記憶部1003に格納された参照モデル1021を用いて、統計量記憶部306cに記憶された標準モデルの統計量を推定する（図4のステップS102c）。なお、この統計量推定部306dによる推定処理は、第3の実施の形態と同様である。

【0371】

次に、初期標準モデル作成部1006bによる初期値の決定方法、つまり、k-means法とマハラノビス汎距離を用いた方法によるクラスタリングについて説明する。図44にクラスタリングのフローチャートを示す。また、図45～図48にクラスタリングのイメージ図を示す。

【0372】

まず、図44のステップS1004において、標準モデルの混合分布数である126個の代表点を準備する（図45）。ここでは、参照モデルの300個の出力分布の中から126個の出力分布を選択して、選択された分布の平均値を代表点とする。

【0373】

次に、図44のステップS1005において、各代表点にマハラノビス汎距離が近い参照モデルの出力ベクトルを決定する（図46）。そいて、図44のステップS1006において、ステップS1005で決定した近い分布を1つのガウス分布で表現して平均値を新しい代表点とする（図47）。

【0374】

続いて、図44のステップS1007において、クラスタリング操作を停止するかどうかを決定する。ここでは、各代表点と参照ベクトルの分布とのマハラノビス汎距離の変化率（1回前の代表点との距離との差分）がしきい値以下になった場合に停止とする。停止

条件を満たさない場合、図44のステップS1005に戻り、近い分布を決定して同様の操作を繰り返す。

【0375】

一方、停止条件を満たす場合には、図44のステップS1008に進み、統計量の初期値を決定して統計量記憶部906cに記憶する。このようにして、クラスタリングによる初期値の決定が行われる。

【0376】

以上説明したように、本発明の第9の実施の形態によれば、アプリケーションに連動して自動的に仕様情報にふさわしい標準モデルを獲得することができる。

なお、標準モデル1022は、音素ごとにHMMを構成してもよい。

【0377】

また、標準モデル作成部1006は、一部の音素の、一部の状態における事象の出力確率に対してモデル作成を行ってもよい。

また、標準モデル1022を構成するHMMは、音素ごとに異なる状態数により構成してもよいし、状態ごとに異なる分布数の混合ガウス分布により構成してもよい。

【0378】

また、標準モデルを作成したのちに、さらに音声データにより学習してもよい。

また、標準モデル構造決定部は、モノフォン、トライフォン、状態共有型などのHMMの構造や、状態数などを決定してもよい。

【0379】

(第10の実施の形態)

図49は、本発明の第10の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。ここでは、本発明に係る標準モデル作成装置がコンピュータシステムにおけるサーバ801に組み込まれた例が示されている。本実施の形態では音声認識用の標準モデル（適応モデル）を作成する場合を例にして説明する。

【0380】

サーバ801は、通信システムにおけるコンピュータ装置等であり、事象の集合と事象または事象間の遷移の出力確率とによって定義される音声認識用の標準モデルを作成する標準モデル作成装置として、読み込み部711と、参照モデル準備部702と、参照モデル記憶部703と、利用情報受信部704と、参照モデル選択部705と、標準モデル作成部706と、仕様情報受信部707と、標準モデル記憶部708と、標準モデル送信部709と、参照モデル受信部810とを備える。

【0381】

参照モデル準備部702は、読み込み部711が読み込んだ、CD-ROMなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部703へ送信する。参照モデル記憶部703は、送信された参照モデル721を記憶する。また、参照モデル準備部702は、端末装置712からの送信に対して参照モデル受信部810が受信した音声認識用参照モデルを参照モデル記憶部703へ送信する。参照モデル記憶部703は、送信された参照モデル721を記憶する。

【0382】

仕様情報受信部707は、端末装置712から仕様情報725を受信する。利用情報受信部704は、端末装置712から利用情報724である雑音下で発声した利用者の音声を受信する。参照モデル選択部705は、利用情報受信部704が受信した利用情報724である利用者の音声に音響的に近い話者・雑音・声調子の参照モデル723を、参照モデル記憶部703が記憶している参照モデル721から選択する。

【0383】

標準モデル作成部706は、仕様情報725に基づいて、参照モデル選択部705が選択した参照モデル723に対する確率または尤度を最大化または極大化するように標準モデル722を作成する処理部であり、第2の実施の形態における標準モデル作成部206と同一の機能を有する。標準モデル記憶部708は、仕様情報725に基づいた1もししく

は複数の標準モデルを記憶する。標準モデル送信部709は、利用者の端末装置712から、仕様情報と標準モデルの要求信号とを受信すると、その端末装置712へ、仕様に適した標準モデルを送信する。

【0384】

次に、以上のように構成されたサーバ801の動作について説明する。

図50は、サーバ801の動作手順を示すフローチャートである。なお、このサーバ801の動作手順を説明するための参照モデルおよび標準モデルの一例は、第7に実施の形態における図31と同様である。

【0385】

まず、標準モデルの作成に先立ち、その基準となる参照モデルを準備する（図50のステップS800、S801）。つまり、参照モデル準備部702は、読み込み部711が読み込んだ、CD-ROMなどのストレージデバイスに書き込まれた話者・雑音・声の調子別の音声認識用参照モデルを参照モデル記憶部703へ送信し、参照モデル記憶部703は、送信された参照モデル721を記憶する（図50のステップS800）。ここでは、参照モデル721は、話者・雑音・声の調子ごとに、音素ごとのHMMにより構成される。また、参照モデル準備部702は、端末装置712が送信して参照モデル受信部810が受信した、利用者と端末装置712に適した音声認識用参照モデルを参照モデル記憶部703へ送信し、参照モデル記憶部703は、送信された参照モデル721を記憶する（図50のステップS801）。ここでは、各参照モデルは、図31の参照モデル721に示されるように、状態数3個、各状態は混合分布数が128個の混合ガウス分布によりHMMの出力分布が構成される。特徴量として25次元（J=25）のメルケプストラム係数が用いられる。

【0386】

以下、これらの参照モデル721を用いた標準モデル722の作成および端末装置712への送信（図50のステップS802～S809）は、第7の実施の形態における手順（図30のステップS701～S708）と同様である。

【0387】

このようにして、端末装置712に記憶された自分用モデルをサーバにアップロードして標準モデル作成の材料にすることができるので、例えば、サーバ801において、アップロードしてきた参照モデルと既に保持している他の参照モデルとを統合して更に混合数の多い高精度の標準モデルを作成し、端末装置712にダウンロードして利用することが可能となる。したがって、端末装置712に簡易的な適応機能が付属され、簡易的に適応したモデルをアップロードして、さらに高精度な標準モデルを作成することもできる。

【0388】

図51は、本実施の形態における標準モデル作成装置を具体的に適用したシステム例を示す図である。ここには、インターネットや無線通信等を介して通信し合うサーバ701と端末装置712（携帯電話機712a、カーナビゲーション装置712b）とが示されている。

【0389】

たとえば、携帯電話機712aは、利用者の音声を利用情報とし、携帯電話機での利用である旨（CPUの処理能力が低いこと）を仕様情報とし、予め記憶しているサンプルモデルを参照モデルとし、それら利用情報、仕様情報および参照モデルをサーバ701に送信することで、標準モデルの作成を要求する。その要求に対してサーバ701で標準モデルが作成されると、携帯電話機712aは、その標準モデルをダウンロードし、その標準モデルを用いて利用者の音声を認識する。例えば、利用者の音声が、内部に保持するアドレス帳の名前と一致した場合には、その名前に対応する電話番号に自動発呼する。

【0390】

また、カーナビゲーション装置712bは、利用者の音声を利用情報とし、カーナビゲーション装置での利用である旨（CPUの処理能力が通常であること）を仕様情報とし、予め記憶しているサンプルモデルを参照モデルとし、それら利用情報、仕様情報および参

照モデルをサーバ701に送信することで、標準モデルの作成を要求する。その要求に対してサーバ701で標準モデルが作成されると、カーナビゲーション装置712bは、その標準モデルをダウンロードし、その標準モデルを用いて利用者の音声を認識する。例えば、利用者の音声が、内部に保持する地名と一致した場合には、その地名を目標点とする現地点からの道順を示す地図を画面に自動表示する。

【0391】

このようにして、携帯電話機712aおよびカーナビゲーション装置712bは、自装置に適した標準モデルの作成をサーバ701に依頼することで、標準モデルの作成に必要な回路や処理プログラムを自装置内に実装する必要がなくなるとともに、様々な認識対象の標準モデルを必要なタイミングで獲得することができる。

【0392】

以上説明したように、本発明の第10の実施の形態によれば、参照モデル受信部810が受信した参照モデルを利用して標準モデルを作成できるため、精度の高い標準モデルが提供される。つまり、端末装置712からのアップロードにより参照モデルを追加することでサーバ801側で保持する参照モデルのバリエーションが増加し、他の人が利用したときにさらに高精度の標準モデルを提供することができる。

【0393】

また、仕様情報に基づいて標準モデルが作成されるため、標準モデルを利用する機器にふさわしい標準モデルが準備される。

【0394】

なお、参照モデル受信部810は、端末装置712とは異なる他の端末装置から参照モデルを受信してもよい。

【0395】

また、図51に示された応用例は、本実施の形態に限られるものではなく、他の実施の形態にも適用することができる。つまり、第1～第9の実施の形態で作成された標準モデルを各種記録媒体や通信を介して様々な電子機器に配信することで、それらの電子機器において、制度の高い音声認識、画像認識、意図理解等を行うことが可能となる。さらに、上記実施の形態における標準モデル作成装置を各種電子機器に内蔵することで、音声認識、画像認識、意図理解等の認識・認証機能を備えるスタンドアローンの電子機器を実現することもできる。

【0396】

以上、本発明に係る標準モデル作成装置について、実施の形態に基づいて説明したが、本発明は、これらの実施の形態に限定されるものではな。

【0397】

たとえば、第1～第10の実施の形態における標準モデルの統計量の近似計算については、各実施の形態における近似計算だけに限らず、第1～第4の実施の形態における合計4種類の近似計算の少なくとも1つを用いてもよい。つまり、4種類の近似計算のいずれであってもよいし、2以上の種類の近似計算の組み合わせであってもよい。

【0398】

また、第2の実施の形態では、統計量推定部206dの一般近似部206eは、標準モデルの混合重み係数、平均値及び分散値を、それぞれ、数45、数46および数47に示される近似式に従って算出したが、これらの近似式に替えて、以下の数63、数64および数65に示される近似式を用いて算出してもよい。

【0399】

【数63】

$$\omega_{f(m)} \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{k=1}^{M_f} \sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, k) v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

$$(m = 1, 2, \dots, M_f)$$

【0400】

【数64】

$$\mu_{f(m,j)} \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} x_{(j)} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

【0401】

【数65】

$$\sigma_{f(m,j)}^2 \approx \frac{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} (x_{(j)} - \mu_{f(m,j)})^2 \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}{\sum_{i=1}^{N_g} \int_{-\infty}^{\infty} \left\{ \sum_{l=1}^{L_{g(i)}} \gamma(\mu_{g(i,l)}, m) v_{g(i,l)} g(x; \mu_{g(i,l)}, \sigma_{g(i,l)}^2) \right\} dx}$$

$$(m = 1, 2, \dots, M_f, j = 1, 2, \dots, J)$$

【0402】

このような近似式を用いて作成した標準モデルによれば、高い認識性能が得られることが発明者らによって確認されている。たとえば、参照モデルと標準モデルそれぞれの混合数を16とした場合の認識結果は、適応前では82.2%であったものが、上記非特許文献2に示された十分統計量による方法では、85.0%、上記近似式による方法では85.5%に改善された。つまり、十分統計量による方法と比べ、高い認識性能が獲得できていることがわかる。また、参照モデルの混合数を64、標準モデルの混合数を16とした場合についての認識結果は、上記近似式による方法では、85.7%と高い認識率が獲得できている。

【0403】

また、初期標準モデル作成部による初期標準モデルの作成においては、図52に示されるようなクラスID・初期標準モデル・参照モデル対応表を予め準備しておき、この表に従って、初期標準モデルを決定してもよい。以下、このようなクラスID・初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定方法について説明する。なお、クラスIDとは、標準モデルを用いた認識対象の種別を識別するIDであり、標準モデルの種類に対応する。

【0404】

図52に示されたクラスID・初期標準モデル・参照モデル対応表は、一定の共通する性質を有する複数の参照モデルに対して、それらを識別する1つのクラスIDを対応づけるとともに、それら参照モデルと共通する性質を持つ予め作成された初期標準モデルを対応づけた表である。この表では、参照モデル8AA～8AZに対して、クラスIDおよび初期標準モデル8Aが対応づけられ、参照モデル64ZA～ZZに対して、クラスIDおよび初期標準モデル64Zが対応づけられている。標準モデル作成部は、使用する参照モデルの性質と共通する初期標準モデルを使用することによって、精度の高い標準モデルを生成することができる。

【0405】

ここで、クラスID、初期標準モデルおよび参照モデルの添え字記号8A、8AAにおける最初の記号「8」等は、混合分布数を意味し、2番目の記号「A」等は大分類、例えば、騒音下における音声認識の場合であれば、騒音環境の種類（家庭内騒音下をA、電車内騒音下をBなど）を意味し、3番目の記号「A」等は小分類、例えば、音声認識の対象となる人の属性（低学年の小学生をA、高学年の小学生をBなど）を意味する。したがって、図52のクラスID・初期標準モデル・参照モデル対応表における参照モデル8AA～AZは、図53に示されるような混合分布数8のモデルであり、参照モデル64ZA～ZZは、図54に示されるような混合分布数64のモデルであり、初期標準モデル8A～64Zは、図55に示されるような混合分布数8～16のモデルである。

【0406】

次に、このようなクラスID・初期標準モデル・参照モデル対応表の作成方法を説明する。図56は、その手順を示すフローチャートであり、図57～図60は、各ステップでの具体例を示す図である。ここでは、騒音環境下での音声認識を例とし、表だけでなく、クラスID、初期標準モデルおよび参照モデルも含めて新規に作成する場合の手順を説明する。

【0407】

まず、音声データを音響的に近いグループに分類する（図56のステップS1100）。たとえば、図57に示されるように、音声データを利用情報である雑音環境で分類する。環境A（家庭内騒音下での音声データ）には、家庭内騒音下で収録した小学生低学年の音声、小学生高学年の音声、成人女性の音声などが含まれ、環境B（電車内での音声データ）には、電車内で収録した小学生低学年の音声、小学生高学年の音声、成人女性の音声などが含まれるように分類する。なお、利用情報である話者の性別、年齢層、笑い声・怒った声などの声の性質、読み上げ調・会話調などの声の調子、英語・中国語などの言語などで分類してもよい。

【0408】

次に、仕様情報等に基づいて、準備する参照モデルの1以上のモデル構造を決定する（図56のステップS1101）。たとえば、8混合、16混合、32混合および64混合を対象とすることを決定する。なお、モデル構造の決定においては、混合分布数を決定するに限らず、HMMの状態数、モノフォン・トライフォンなどのHMMの種類などを決定してもよい。

【0409】

続いて、初期標準モデルを作成する（図56のステップS1102）。つまり、上記音声データの分類（ステップS1100）において決定した分類（環境A、環境B、…）ご

とに、ステップS1101において決定したモデル構造ごとの初期標準モデルを作成する。例えば、図58に示されるように、初期標準モデル8Aであれば、8混合の初期標準モデルを、家庭内騒音下（環境A）における音声データ（低学年的小学生、高学年的小学生、成人男、成人女等の音声データ）を用いて、バウム・ウェルチアルゴリズムなどにより学習して作成する。

【0410】

次に、参照モデルを作成する（図56のステップS1103）。つまり、上記ステップS1102において作成した初期標準モデルを用いて参照モデルを作成する。具体的には、参照モデルを学習する音声データの雑音環境と同じ雑音環境で学習した、同じ混合分布数をもつ初期標準モデルを用いて参照モデルを学習する。例えば、図59に示されるように、参照モデル8AAは、混合分布数8の家庭内騒音下での小学生低学年の音声データで学習するモデルであり、学習を行う際の初期値として、同じ環境である家庭内騒音下での音声データ（小学生低学年、小学生高学年、成人女性、成人男性の音声を含む）で学習した初期標準モデルを用いる。学習方法として、バウム・ウェルチアルゴリズムを用いる。

【0411】

最後に、クラスIDを付与する（図56のステップS1104）。たとえば、騒音環境下ごとに1つのクラスIDを付与することによって、図60に示されるクラスID・初期標準モデル・参照モデル対応表、つまり、“クラスID付き初期標準モデル”および“クラスID付き参照モデル”が作成される。

【0412】

なお、このようなクラスID・初期標準モデル・参照モデル対応表は、完成された表として予め端末（標準モデル作成装置）が保持している必要はない。端末（標準モデル作成装置）は、図61に示されるように、他の装置（サーバ）と通信することによって表を完成させてもよい。つまり、標準モデル作成装置（端末）は、通信網などを介して、“クラスID付き初期標準モデル”，“クラスID付き参照モデル”を取得することが可能である。もっとも、端末は必ずしも“クラスID付き初期標準モデル”，“クラスID付参照モデル”を取得する必要はなく事前に記憶させて出荷してもよい。

【0413】

図61に示されるように、端末は、以下のようない方法によって、“クラスID付き初期標準モデル”、“クラスID付き参照モデル”を取得することができる。第1の方法として、端末は、“クラスID付き初期標準モデル”（例えば規格化コンソーシアムなどで事前に定義されたクラスIDのつけ方に遵守したもの）を記憶しているケースである。このとき、端末は、1以上のサーバから“クラスID付き参照モデル”（例えば規格化コンソーシアムなどで事前に定義されたクラスIDのつけ方に遵守したもの）をダウンロードする。なお、端末に、“クラスID付き参照モデル”を出荷時に記憶させておいてもよい。

【0414】

また、第2の方法として、端末は、“クラスID付き初期標準モデル”を記憶していないケースである。このとき、端末は、サーバ（図61のサーバ1）から“クラスID付き初期標準モデル”をダウンロードする。次に、端末は、1以上のサーバ（図61のサーバ2）から“クラスID付き参照モデル”をダウンロードする。必要に応じて逐次的にクラスIDの定義の追加、変更が可能である。また、端末のメモリの節約にもなる。

【0415】

さらに、第3の方法として、端末は、クラスIDと初期標準モデル・参照モデルの対応関係を明記した“クラスID・初期標準モデル・参照モデル対応表”を記憶しているケースである。このとき、端末は、“対応表”記憶していないサーバ（図61のサーバ3）に“対応表”をアップロードする。サーバは、送信された“対応表”に基づき“クラスID付き参照モデル”を準備する。端末は、準備された“クラスID付き参照モデル”をダウンロードする。

【0416】

次に、このようなクラスID・初期標準モデル・参照モデル対応表を用いた初期標準モ

デル作成部による初期標準モデルの決定方法について説明する。図62は、その手順を示すフローチャートである。図63および図64は、各ステップでの具体例を示す図である。

【0417】

まず、標準モデルの作成に用いる参照モデルからクラスIDを抽出する（図62のステップS1105）。たとえば、図63に示されるテーブルに従って、選択された参照モデルから、対応するクラスIDを抽出する。ここでは、抽出したクラスIDとして、8Aが1個、16Aが3個、16Bが1個、64Bが1個とする。

【0418】

次に、抽出したクラスIDを用いて標準モデル作成に用いる初期標準モデルを決定する（図62のステップS1106）。具体的には、以下の手順に従って初期標準モデルを決定する。

(1) 作成する標準モデルの混合分布数(16混合)と同じクラスID(16*)をもつ参照モデルから抽出したクラスID(16A、16B)に着目し、その中から一番多く抽出されたクラスIDに対応する初期標準モデルを最終的な初期標準モデルと決定する。たとえば、標準モデルの構造が16混合の場合には、16混合に関するクラスIDとして、16Aが3個、16Bが1個抽出されているので、クラスIDが16Bの初期標準モデルを採用する。

(2) 作成する標準モデルの混合分布数(8混合)と同じクラスID(8*)をもつ参照モデルから抽出したクラスID(8A)に着目し、同じクラスIDをもつ初期標準モデルを最終的な初期標準モデルと決定する。たとえば、標準モデルの構造が8混合の場合には、8混合に関するクラスIDとして、8Aが1個抽出されているので、クラスIDが8Aの初期標準モデルを採用する。

(3) 作成する標準モデルの混合分布数(32混合)と同じクラスID(32*)をもつ参照モデルから抽出したクラスIDに着目し、存在しない場合、仕様情報に着目してその中から一番多く抽出されたクラスID(*A)をもつ初期標準モデル(8A、16A)を用いてクラスタリングにより32混合にして最終的な初期標準モデルとする（図44を参考）。たとえば、標準モデルの構造が32混合の場合には、32混合に関するクラスIDが抽出されていないので、一番多く抽出されたクラスID(16A)を用いてクラスタリングにより32混合にして初期標準モデルとする。

【0419】

なお、はじめに作成する標準モデルの仕様情報（混合分布数など）に着目せず、利用情報（雑音の種類など）に着目して初期値を決定してもよい。

【0420】

図64に、第3近似部を用いて作成した混合分布数が64の標準モデルを用いた認識実験の結果を示す。縦軸に成人（男性と女性）の認識率（%）、横軸に繰り返し回数Rを示す。繰り返し回数R=0とは、学習を行う前での初期標準モデル作成部が作成した初期モデルにより認識した結果である。また、繰り返し回数R=1～5において、近傍指示パラメータG=1とした。

【0421】

グラフ「データ」は、数日間かけて音声データより学習した場合の結果を表しており、グラフ「女性」、グラフ「男性」は、それぞれ、初期モデルを成人女性、成人男性としたときの結果を表している。参照モデルによる本発明による学習時間は数分のオーダーであった。この実験結果より、成人女性の参照モデルを初期標準モデルと決定した場合には、音声データで学習した結果よりも高い精度の標準モデルが作成できていることが分かる。

【0422】

このことは、音声データを分割し、分割した音声データをそれぞれの参照モデルとして厳密に学習したのちに統合したほうが、音声データによる学習の課題である局所解に陥るという問題を解決できる可能性を示している（音声データによる学習との認識精度での比較）。

【0423】

また、音声データの収録が困難な子供の音声データに対しては、データ数に適切である混合分布数の少ない参照モデルで厳密に学習して、多くの音声データの収録が可能な成人の音声データに対しては、混合分布数の多い参照モデルで厳密に学習して、その後で本発明により統合して標準モデルを作成すれば、極めて精度の高い標準モデルが作成できることが期待できる。

【0424】

なお、標準モデルの混合分布数が16の場合における認識実験（図39）では、本発明による方法は、音声データで学習した標準モデルの認識率を超えていない。このことは、音声データを16混合の参照モデルの形にしたときに音声データの情報が欠如したためだと考えられる。参照モデルを64混合で作成して音声データの特徴を十分保持しておけばより高い精度の標準モデルが作成できる。のことより、第9の実施の形態では、参照モデルの混合分布数を300と大きめに設定している。

【0425】

また、図39および図64に示される認識実験より、初期標準モデルが認識精度に与える影響が示されており、初期標準モデルの決定方法の重要性を物語っている（図64において、成人女性の参照モデルを初期標準モデルとして利用した場合、成人男性の参照モデルを利用する場合より高い精度の標準モデルが作成できることが示されている）。

【0426】

以上のように、クラスID・初期標準モデル・参照モデル対応表に従って、参照モデルと共に性質の初期標準モデルを用いることで、精度の高い標準モデルを作成することができる。

【0427】

なお、このようなクラスID・初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定は、上記実施の形態1～10のいずれにおいても採用することができる。
【産業上の利用可能性】

【0428】

本発明に係る標準モデル作成装置は、確率モデル等を用いた音声、文字、画像等の対象物を認識する装置等として利用することができ、例えば、音声によって各種処理を実行するテレビ受信装置・カーナビゲーション装置、音声を他の言語に翻訳する翻訳装置、音声で操作するゲーム装置、音声による検索キーワードで情報を検索する検索装置、人物検出・指紋認証・顔認証・虹彩認証等を行う認証装置、株価予測、天気予測などの予測を行う情報処理装置等として利用することができる。

【図面の簡単な説明】

【0429】

【図1】本発明の第1の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

【図2】同サーバの動作手順を示すフローチャートである。

【図3】図1における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

【図4】図2におけるステップS101（標準モデルの作成）の詳細な手順を示すフローチャートである。

【図5】図1における第1近似部104eによる近似計算を説明する図である。

【図6】参照モデルを選択する際の画面表示例を示す図である。

【図7】(a)は、作成する標準モデルの構造（混合分布数）を指定する際の画面表示例を示し、(b)は、仕様情報を選択する際の画面表示例を示す図である。

【図8】標準モデルを作成しているときの進歩状況を示す画面表示例を示す図である。

【図9】本発明の第2の実施の形態における標準モデル作成装置に係るSTBの全体構成を示すブロック図である。

【図10】同STBの動作手順を示すフローチャートである。

【図11】図10における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

【図12】図10における第2近似部による近似計算を説明する図である。

【図13】本発明の第3の実施の形態における標準モデル作成装置に係るPDAの全体構成を示すブロック図である。

【図14】同PDAの動作手順を示すフローチャートである。

【図15】図13における参照モデル記憶部に記憶されている参照モデルの例を示す図である。

【図16】同PDAの選択画面の一例を示す。

【図17】図13における統計量推定部による統計量の推定手順を示す概念図である。

【図18】図13における第3近似部による近似計算を説明する図である。

【図19】本発明の第4の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

【図20】同サーバの動作手順を示すフローチャートである。

【図21】同サーバの動作手順を説明するための参照モデルおよび標準モデルの一例を示す図である。

【図22】利用情報としての個人情報を入力する際の画面表示例を示す図である。

【図23】本発明の第5の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

【図24】同サーバの動作手順を示すフローチャートである。

【図25】同サーバの動作手順を説明するための参照モデルおよび標準モデルの一例を示す図である。

【図26】本発明の第6の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

【図27】同サーバの動作手順を示すフローチャートである。

【図28】同サーバの動作手順を説明するための参照モデルおよび標準モデルの一例を示す図である。

【図29】本発明の第7の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

【図30】同サーバの動作手順を示すフローチャートである。

【図31】同サーバの動作手順を説明するための参照モデルおよび標準モデルの一例を示す図である。

【図32】本発明の第8の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。

【図33】携帯電話901の動作手順を示すフローチャートである。

【図34】参照モデル記憶部に格納されている参照モデルの一例を示す図である。

【図35】新たに参照モデル記憶部に格納された参照モデルの一例を示す図である。

【図36】利用情報を作成するときの画面表示例を示す図である。

【図37】参照モデルを準備するときの画面表示例を示す図である。

【図38】第3近似部を用いて作成した標準モデルを用いた認識実験の結果を示すグラフである。

【図39】第3の実施の形態における第2近似部により作成された標準モデルによる認識実験の結果を示すグラフである。

【図40】本発明の第9の実施の形態における標準モデル作成装置の全体構成を示すブロック図である。

【図41】アプリ・仕様情報対応データベースのデータ例を示す図である。

【図42】PDA1001の動作手順を示すフローチャートである。

【図43】参照モデル記憶部に格納されている参照モデルの一例を示す図である。

【図44】初期標準モデル作成部によるクラスタリングによる初期値の決定方法を示すフローチャートである。

【図45】図44におけるステップS1004の具体例を示す図である。

【図46】図44におけるステップS1005の具体例を示す図である。

【図47】図44におけるステップS1006の具体例を示す図である。

【図48】図44におけるステップS1008の具体例を示す図である。

【図49】本発明の第10の実施の形態における標準モデル作成装置に係るサーバの全体構成を示すブロック図である。

【図50】同サーバの動作手順を示すフローチャートである。

【図51】本発明に係る標準モデル作成装置を具体的に適用したシステム例を示す図である。

【図52】クラスID・初期標準モデル・参照モデル対応表の例を示す図である。

【図53】図52のクラスID・初期標準モデル・参照モデル対応表における参照モデル8AA～AZの例を示す図である。

【図54】図52のクラスID・初期標準モデル・参照モデル対応表における参照モデル64ZA～ZZの例を示す図である。

【図55】図52のクラスID・初期標準モデル・参照モデル対応表における初期標準モデル8A～64Zの例を示す図である。

【図56】クラスID・初期標準モデル・参照モデル対応表の作成方法を示すフローチャートである。

【図57】図56におけるステップS1100の具体例を示す図である。

【図58】図56におけるステップS1102の具体例を示す図である。

【図59】図56におけるステップS1103の具体例を示す図である。

【図60】図56におけるステップS1104の具体例を示す図である。

【図61】端末がサーバと通信することによってクラスID・初期標準モデル・参照モデル対応表を完成させる手順を示す図である。

【図62】クラスID・初期標準モデル・参照モデル対応表を用いた初期標準モデルの決定方法を示すフローチャートである。

【図63】図62におけるステップS1105の具体例を示す図である。

【図64】第3近似部を用いて作成した標準モデルを用いた認識実験の結果を示すグラフである。

【符号の説明】

【0430】

101、401、501、601、701、801 サーバ

102、202、302、402、502、602、702、902 参照モデル

準備部

103、203、303、403、503、603、703、903、1003

参照モデル記憶部

104、206、306、406、506、606、706、906、1006

標準モデル作成部

104a、206a、306a、906a、1006a 標準モデル構造決定部

104b、206b、306b、906b、1006b 初期標準モデル作成部

104c、206c、306c、906c 統計量記憶部

104d、206d、306d、906d 統計量推定部

104e 第1近似部

111、311、511、611、711 読み込み部

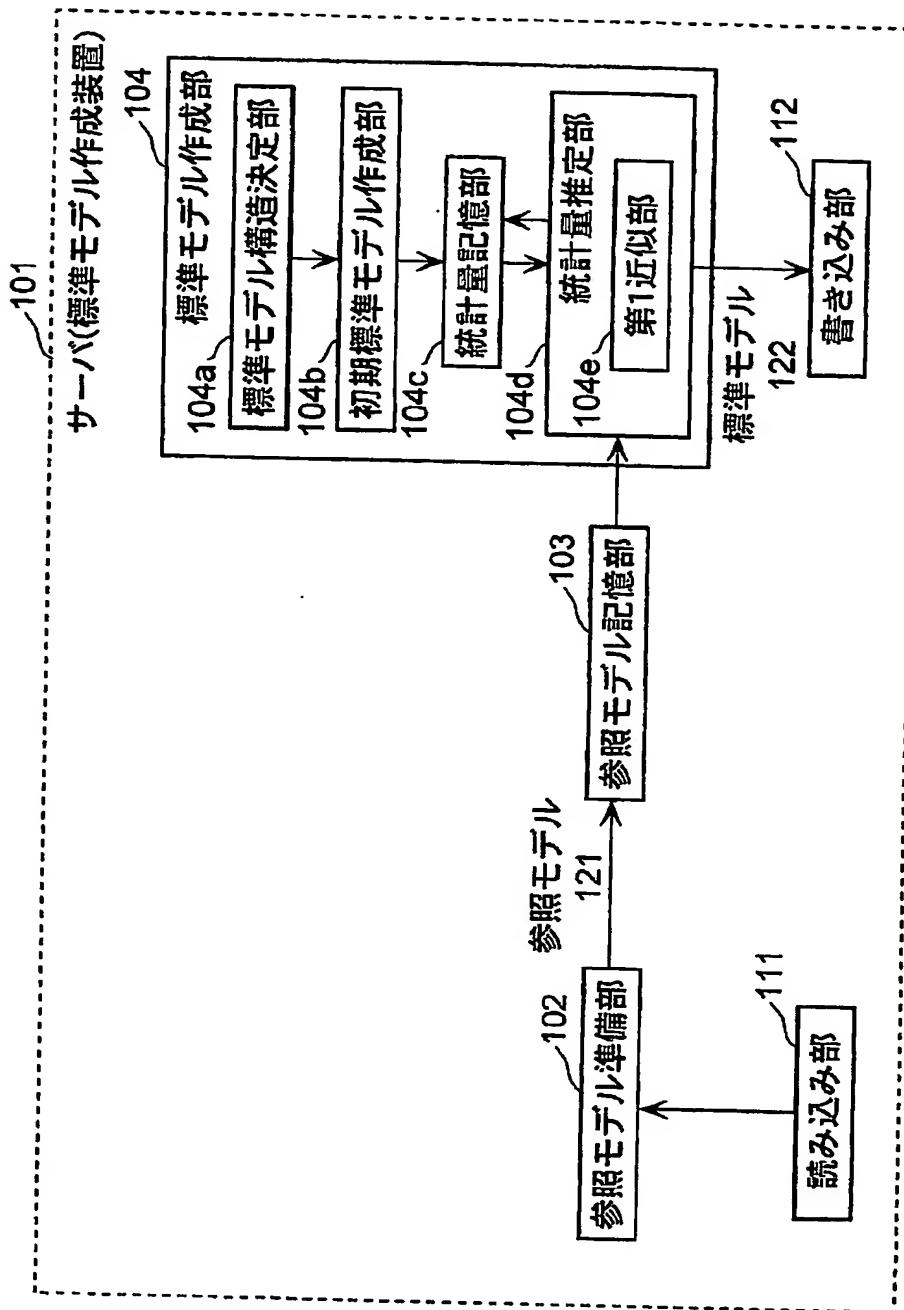
112、413、513 書き込み部

201 STB

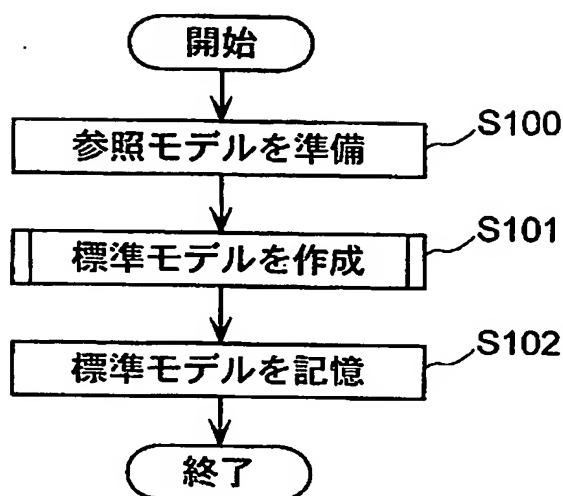
204、304、904 利用情報作成部

205、305、405、505、605、705、905 参照モデル選択部
206e 一般近似部
211、312、912、1012 マイク
212、512 音声データ蓄積部
213、913、1013 音声認識部
301、1001 PDA
306e 第2近似部
307、607、907 仕様情報作成部
313 雜音識別部
404、504、604、704 利用情報受信部
411 カメラ
412 画像データ蓄積部
414 電話
507、707 仕様情報受信部
512、614、712 端末装置
708 標準モデル記憶部
709 標準モデル送信部
712a、901 携帯電話機
712b カーナビゲーション装置
810、909 参照モデル受信部
906e 第3近似部
908 類似度情報作成部
910 サーバ装置
1014 アプリ・仕様情報対応データベース

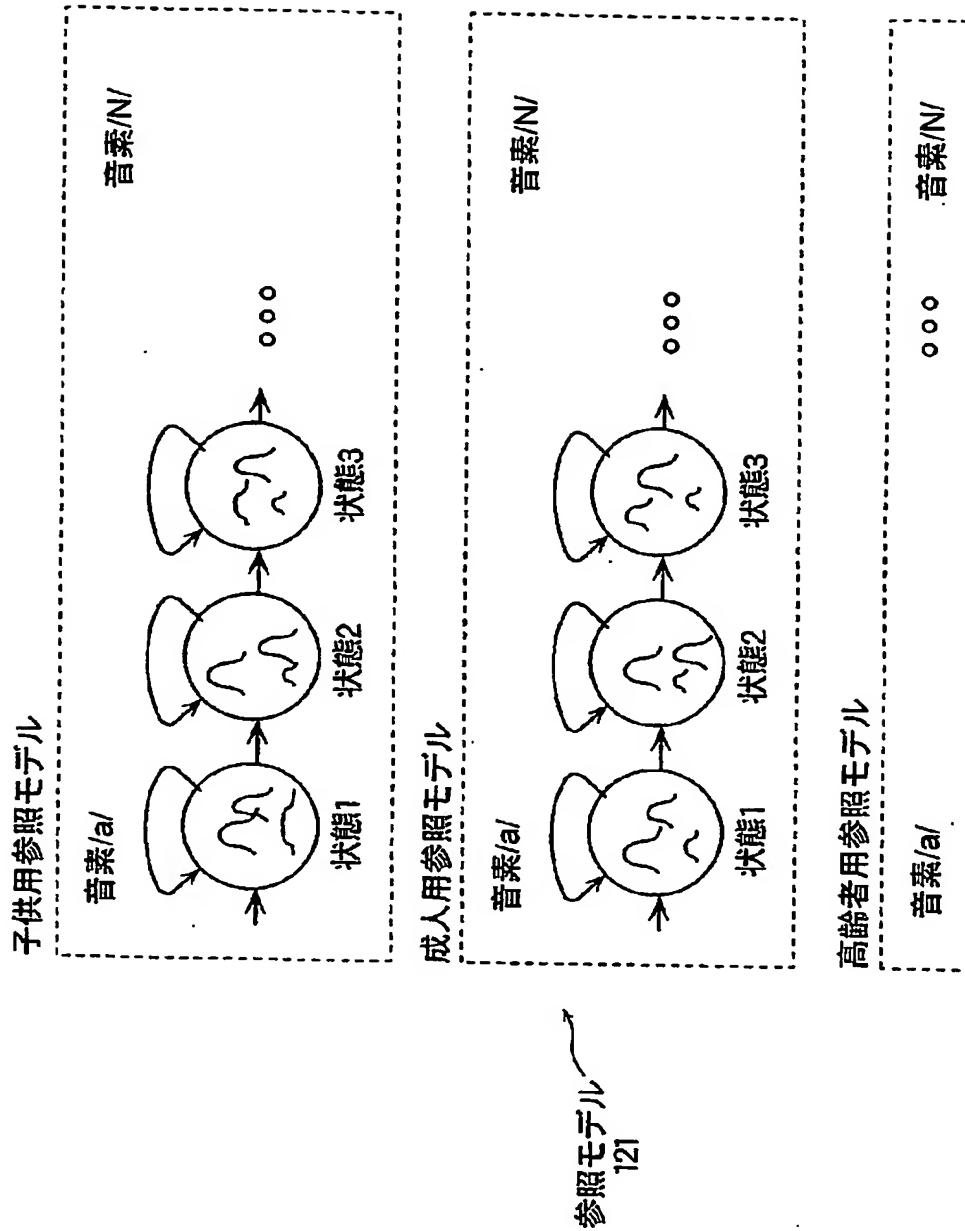
【書類名】 図面
【図 1】



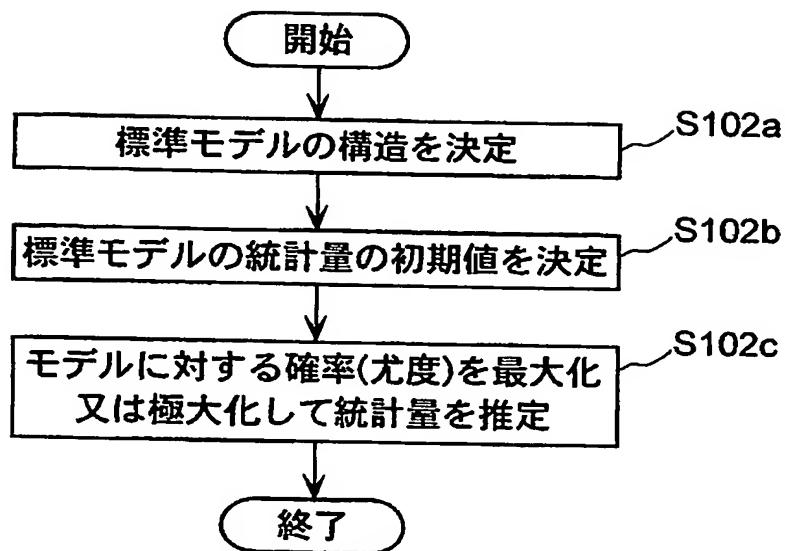
【図2】



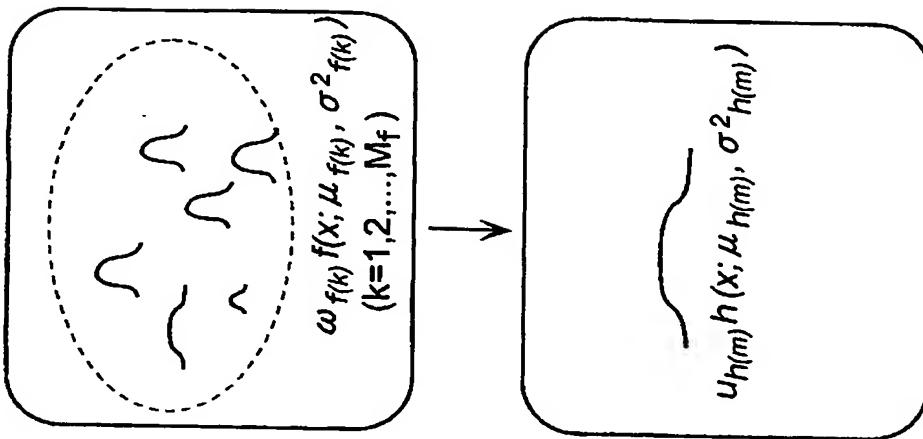
【図3】



【図4】



【図5】



(b)

近似式

$$\gamma(x, m) = \frac{\omega_{f(m)} f(x; \mu_{f(m)}, \sigma^2_{f(m)})}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x; \mu_{f(k)}, \sigma^2_{f(k)})} \approx \frac{\omega_{f(m)} f(x; \mu_{f(m)}, \sigma^2_{f(m)})}{\omega_{h(m)} h(x; \mu_{h(m)}, \sigma^2_{h(m)})}$$

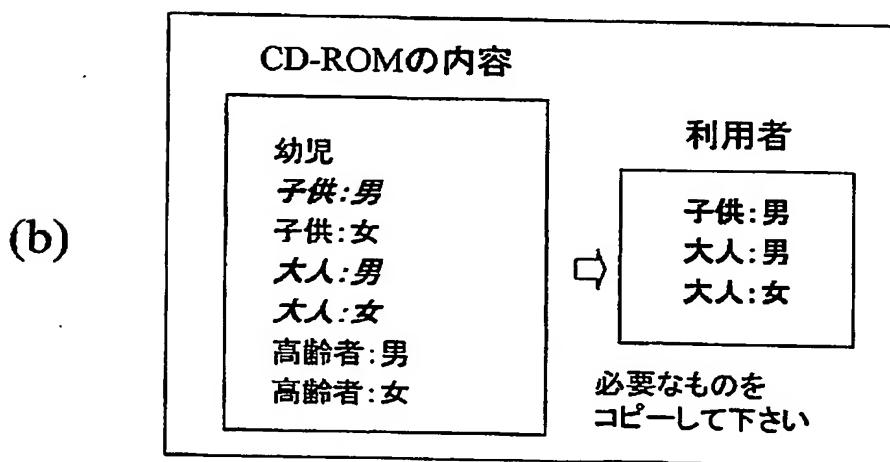
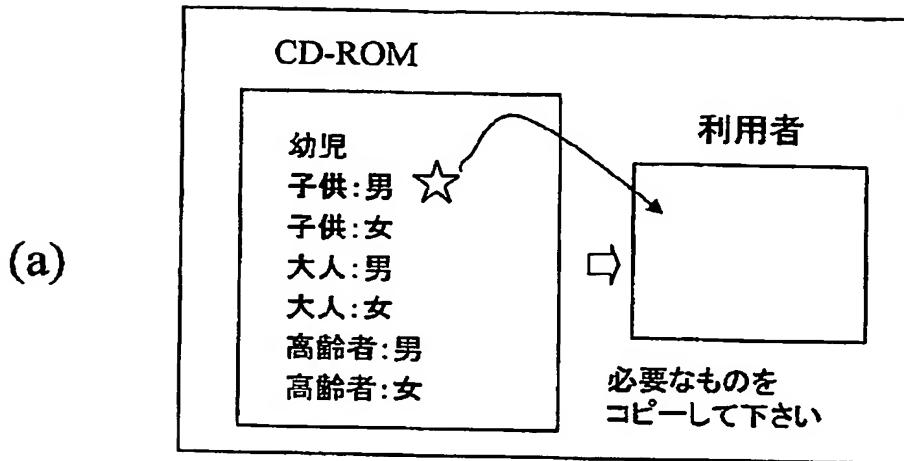
$\omega_{f(k)} f(x; \mu_{f(k)}, \sigma^2_{f(k)})$
($k=1, 2, \dots, M_f$)

$u_{h(m)} h(x; \mu_{h(m)}, \sigma^2_{h(m)})$

を全ての
 $\omega_{f(k)} f(x; \mu_{f(k)}, \sigma^2_{f(k)})$
($k=1, 2, \dots, M_f$)
で作成する

(a)

【図6】



【図7】

(a)

モデルの複雑さ(混合分布数)

- ① 3個
- ② 10個
- ③ 20個

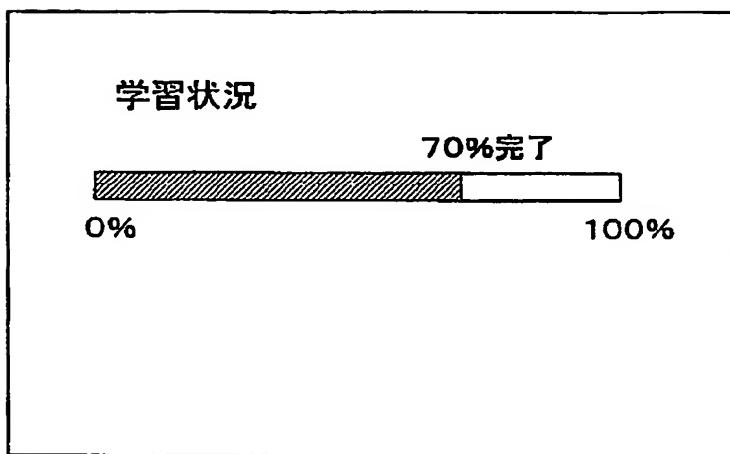
(b)

利用機器

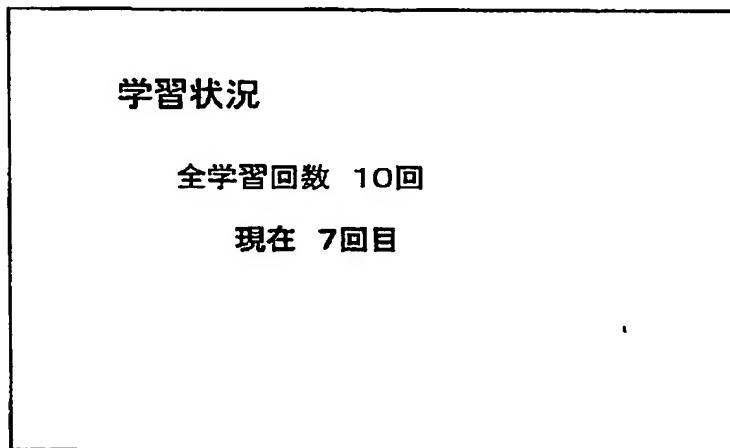
- ① テレビ用
- ② カーナビ用
- ③ 携帯電話用

【図8】

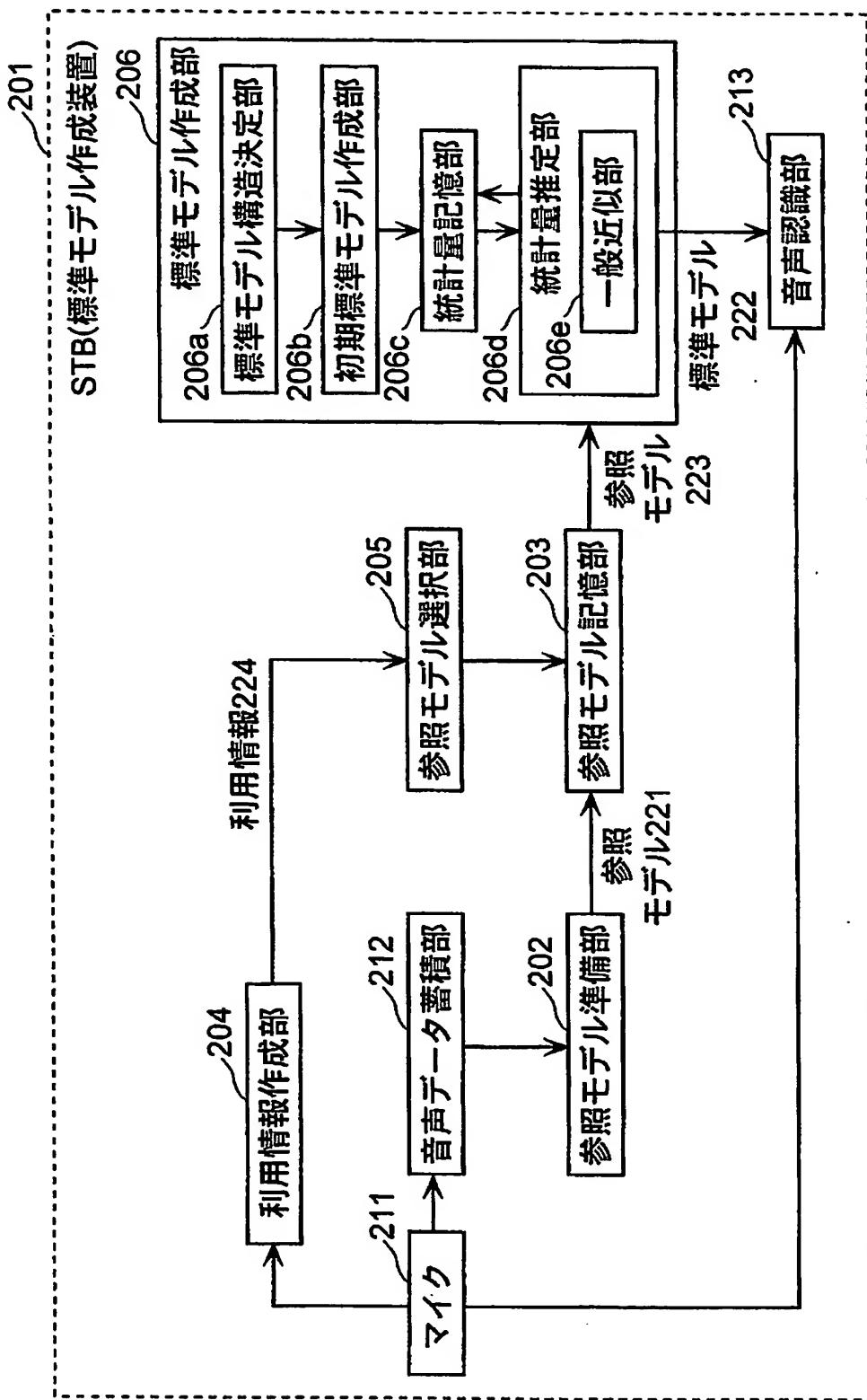
(a)



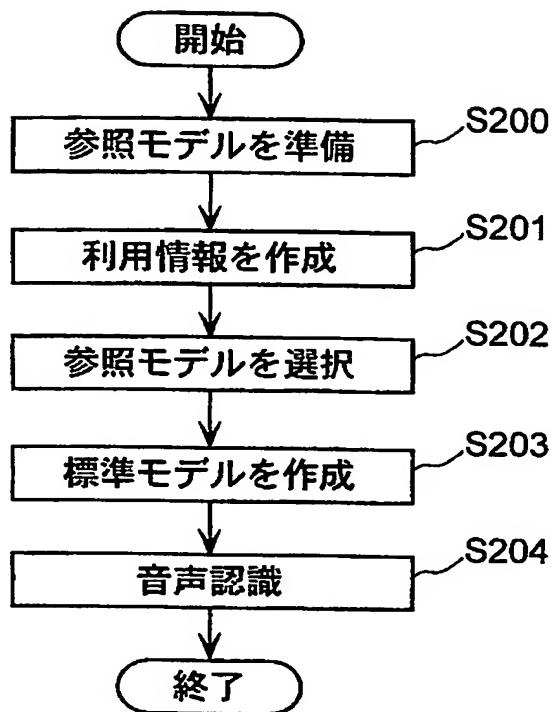
(b)



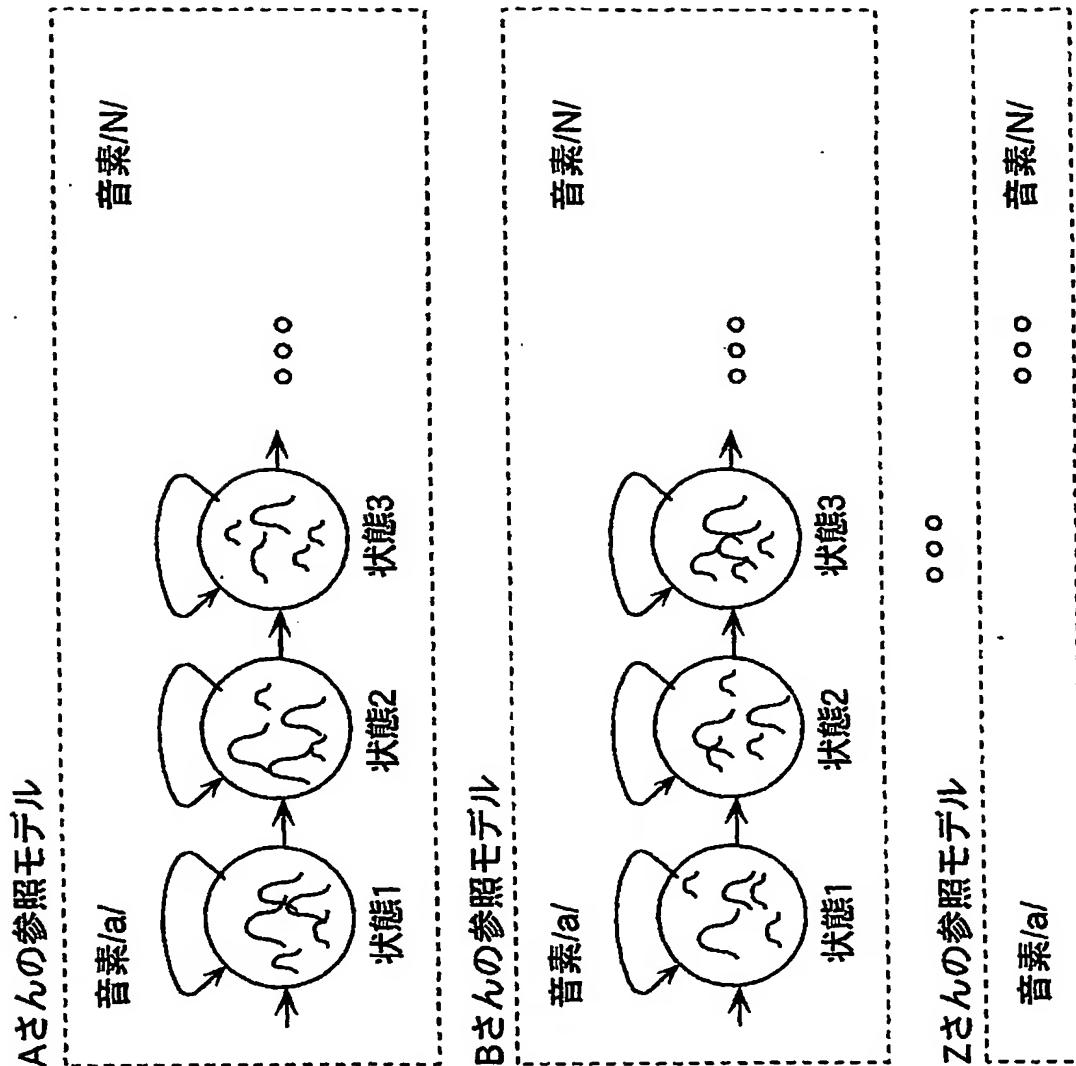
〔図9〕



【図10】



【図11】



参考モデル
221

【図12】

近似式

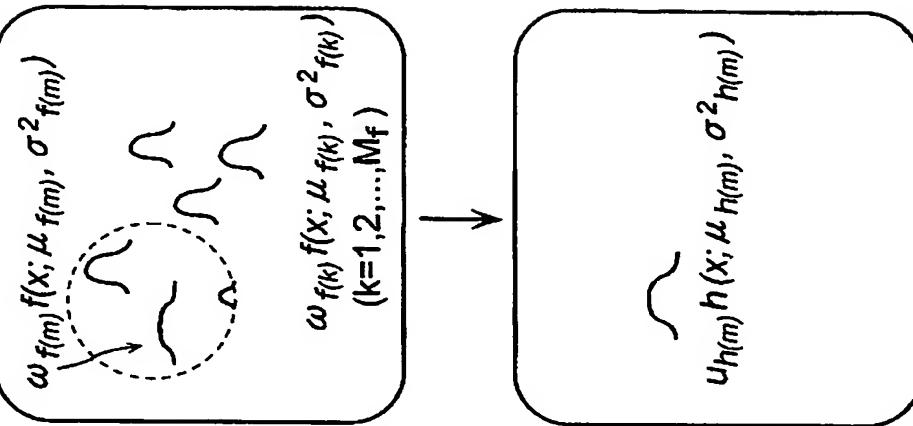
$$\gamma(x, m) = \frac{\omega_{f(m)} f(x; \mu_{f(m)}, \sigma^2_{f(m)})}{\sum_{k=1}^{M_f} \omega_{f(k)} f(x; \mu_{f(k)}, \sigma^2_{f(k)})} \approx \frac{\omega_{f(m)} f(x; \mu_{f(m)}, \sigma^2_{f(m)})}{u_{h(m)} h(x; \mu_{h(m)}, \sigma^2_{h(m)})}$$

$u_{h(m)} h(x; \mu_{h(m)}, \sigma^2_{h(m)})$
を

$\omega_{f(m)} f(x; \mu_{f(m)}, \sigma^2_{f(m)})$
に距離が近い

$\omega_{f(k)} f(x; \mu_{f(k)}, \sigma^2_{f(k)})$
($k=1, 2, \dots, M_f$)
 $P_{h(m)}$ は1以上
で作成する

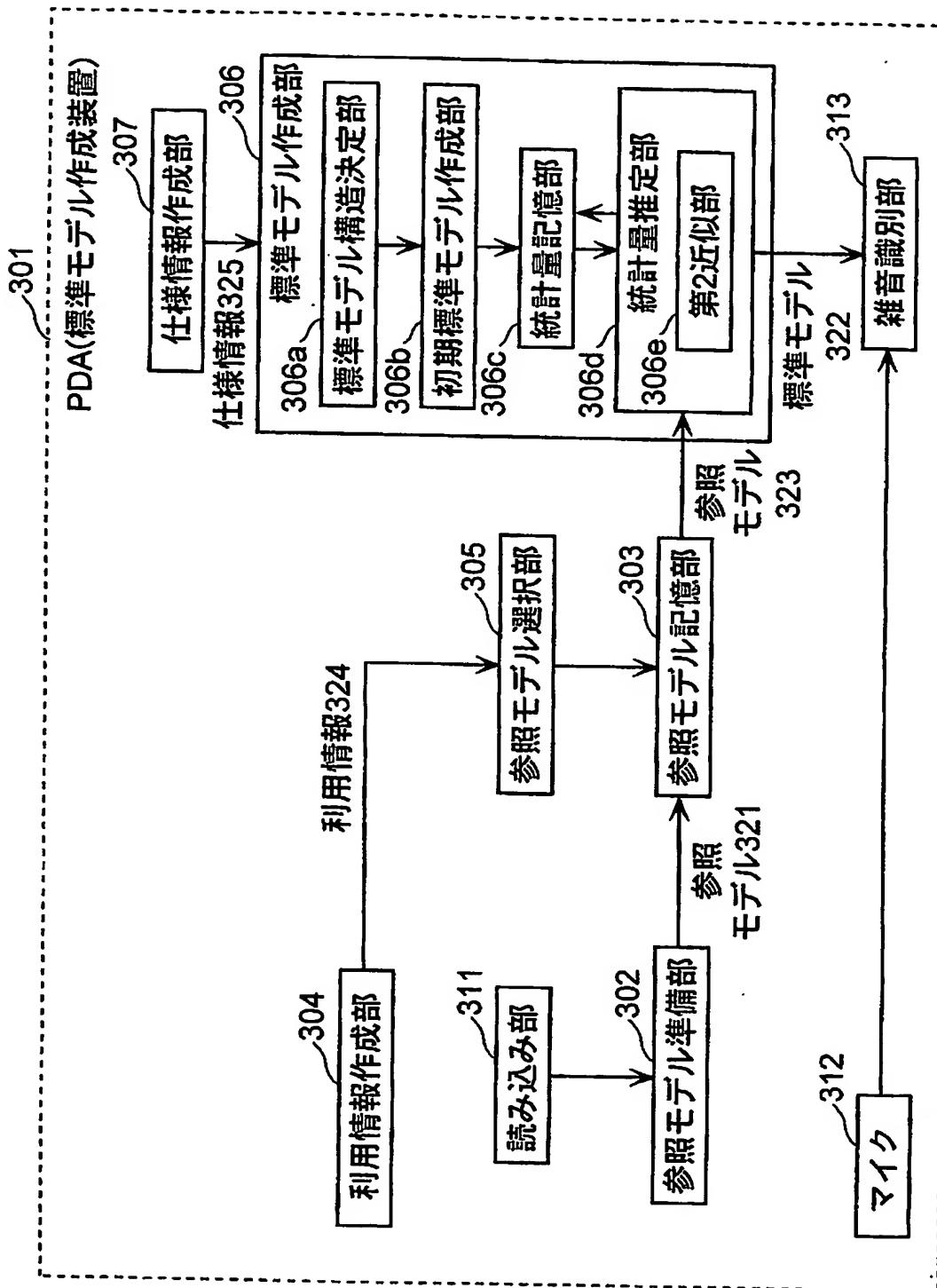
距離：平均値のユークリッド距離
マハラノビス距離
カルバック・ライブラー(KL)距離 など



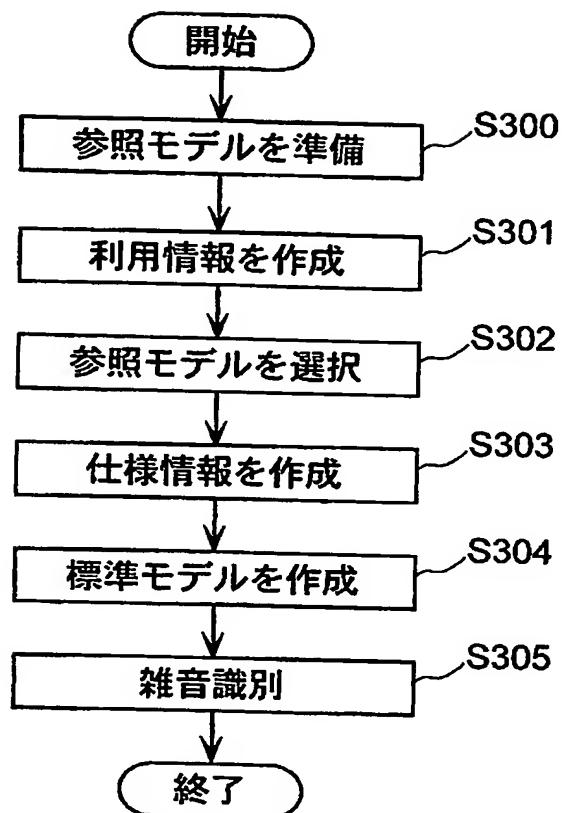
(a)

(b)

【図13】

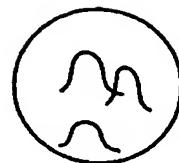


【図14】



【図15】

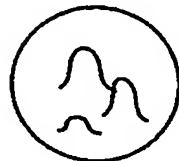
乗用車Aの参考モデル



乗用車Bの参考モデル



バスAの参考モデル



参考モデル321 →

小雨の参考モデル



大雨の参考モデル



【図16】

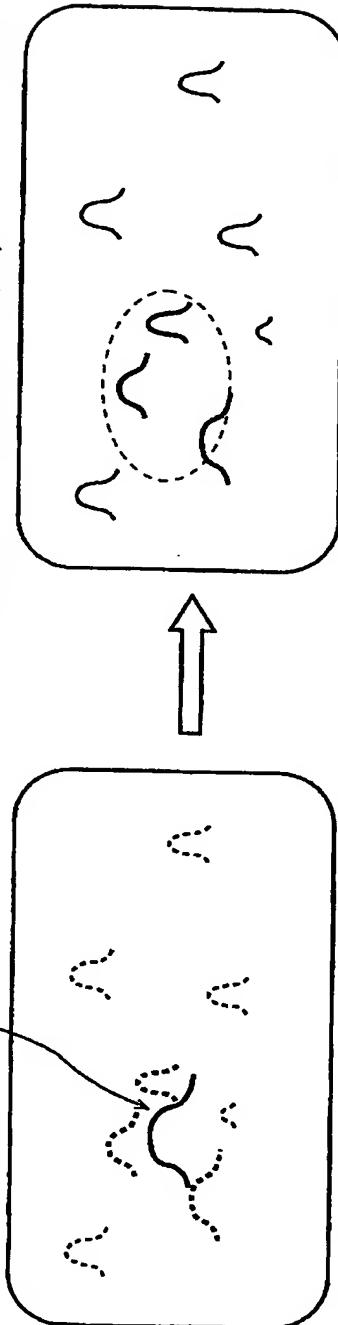
識別する雑音の種類

- 1. 自動車**
 - 1.1. 乗用車**
 - 1.2. バス
 - 1.3. トラック
- 2. 雨
- 3. 飛行機
- 4. 警報音

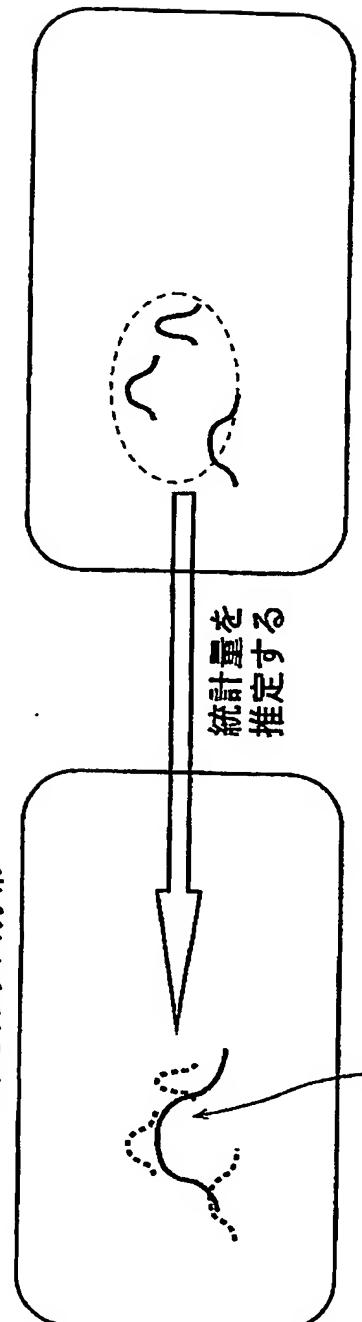
【図17】

標準モデルのガウス分布($m=1, 2, \dots, M_f$) 各モデルの各参照ガウス分布

$$\omega_{f(m)} f(x; \mu_{f(m)}, \sigma^2_{f(m)}) \quad (\text{推定前}) \quad \omega_{g(i, \theta)} g(x; \mu_{g(i, \theta)}, \sigma^2_{g(i, \theta)}) \quad (i=1, 2, \dots, N_g, \theta=1, 2, \dots, L_{(i)})$$

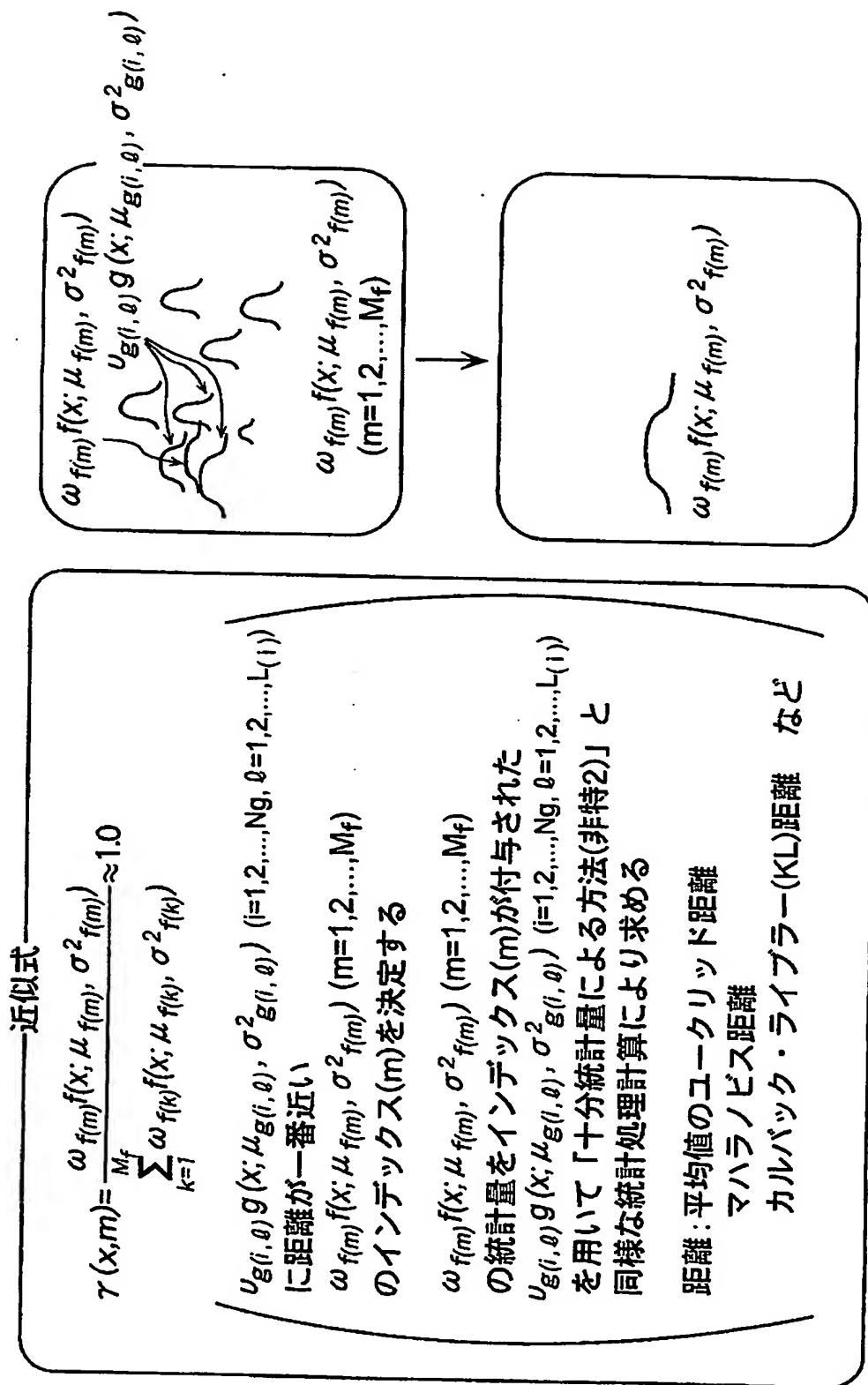


各モデルの各ガウス分布に対して、マハラノビス距離が最も近いものが標準モデルのガウス分布であるガウス分布



$$\omega_{f(m)} f(x; \mu_{f(m)}, \sigma^2_{f(m)}) \quad (\text{推定後})$$

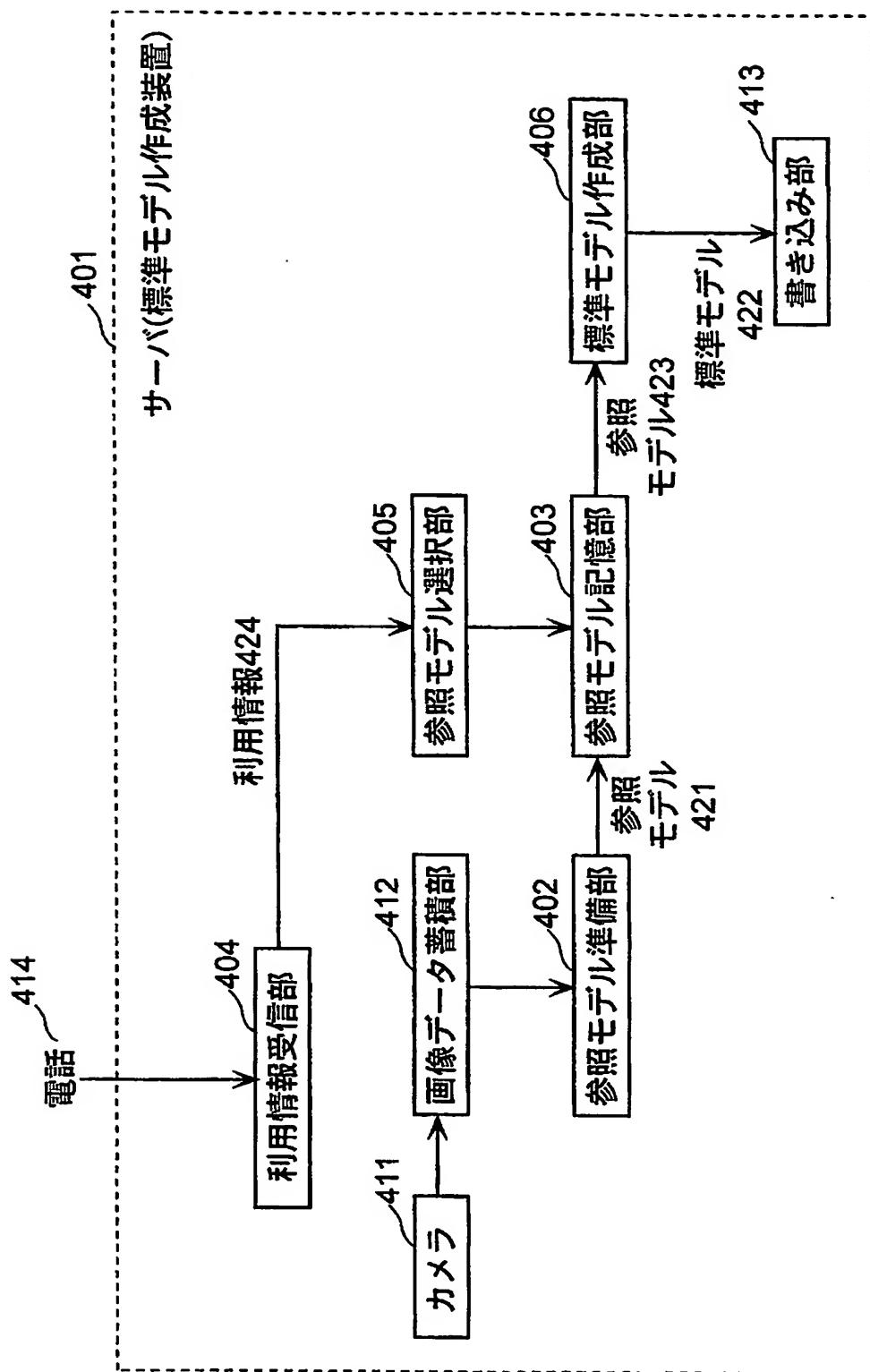
【図18】



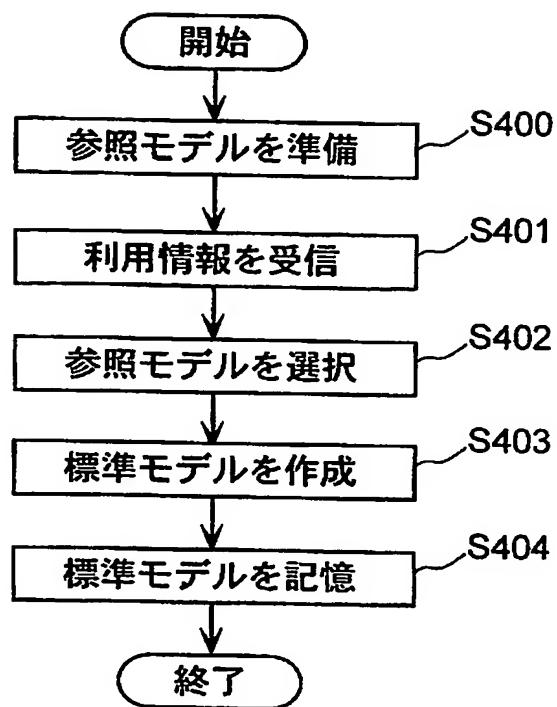
(a)

(b)

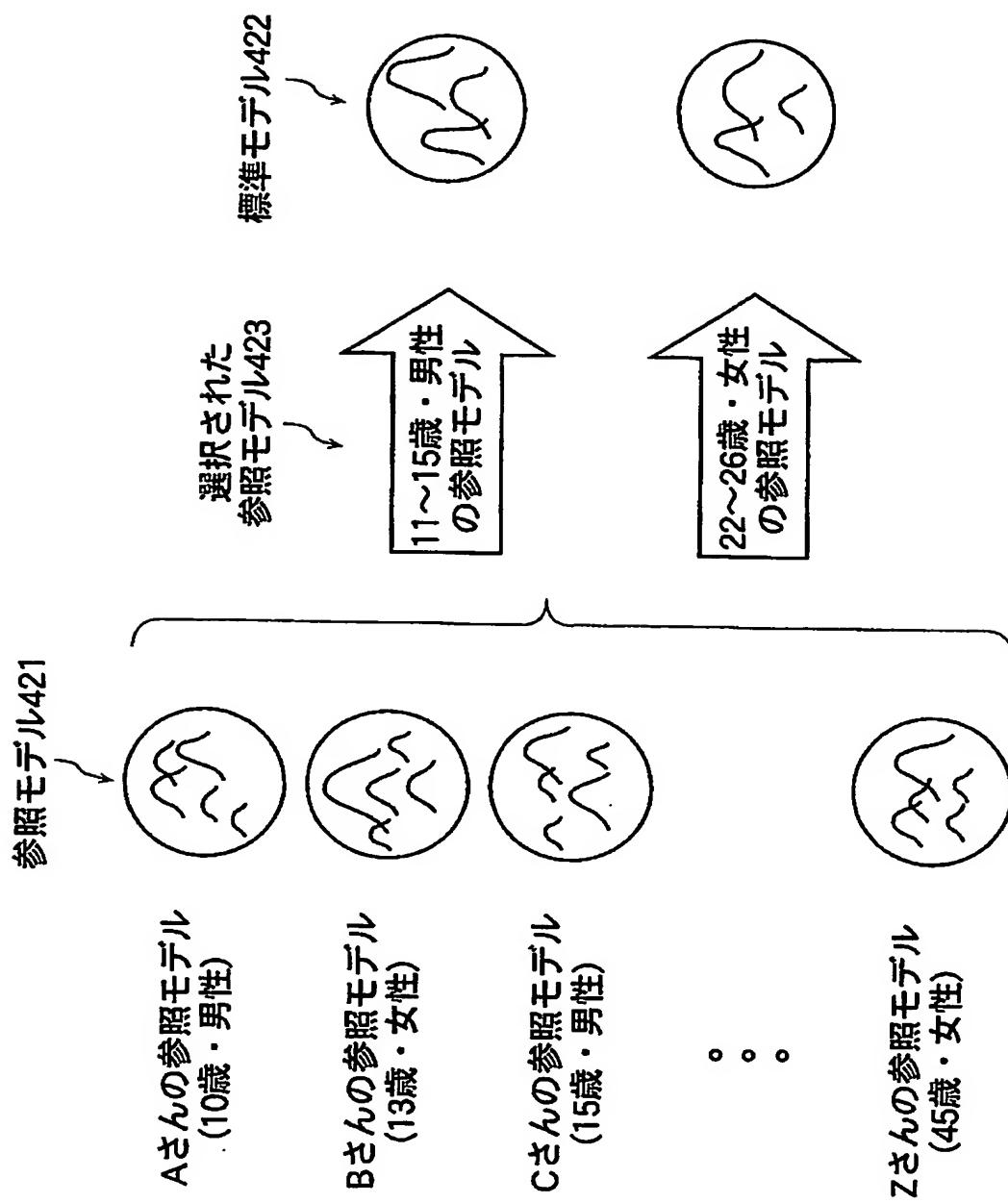
【図19】



【図20】



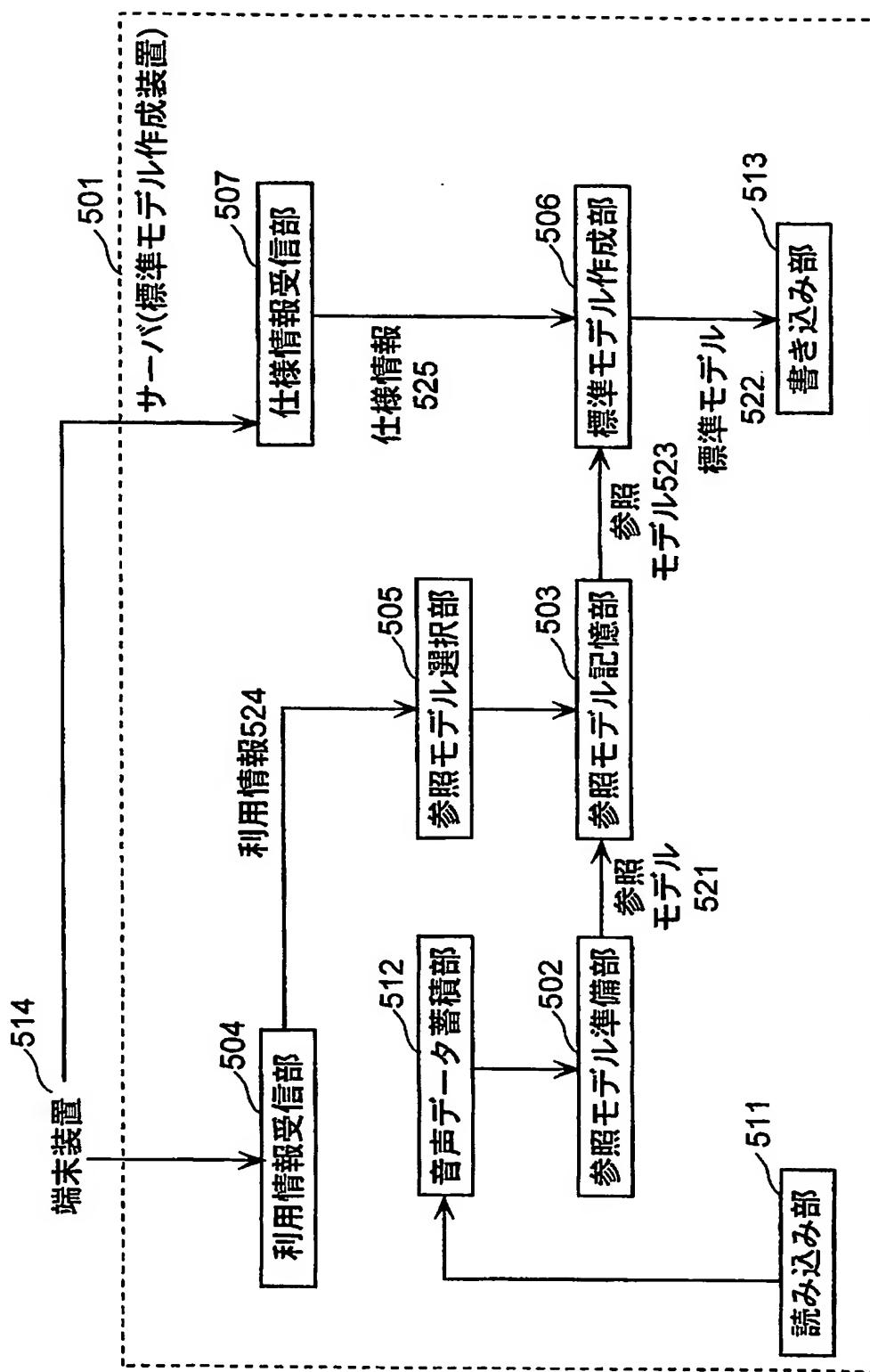
【図21】



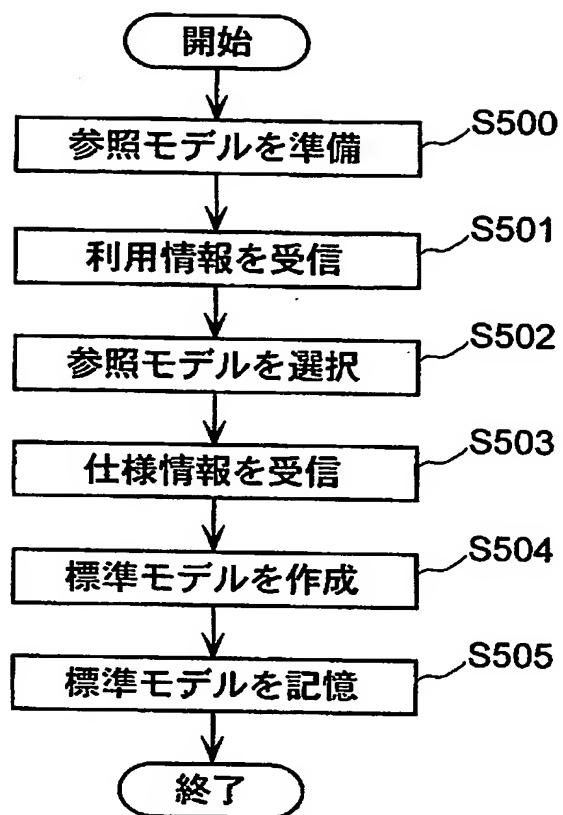
【図22】

名前 「お父さん」	性別 ■男 □女
住所 「大阪市」	年齢 「50歳」
趣味 ■ドライブ □スポーツ観戦 □つり □ショッピング ■温泉 □ゴルフ □キヤンブ	
□マッサージ □コンピュータ □ゲーム	

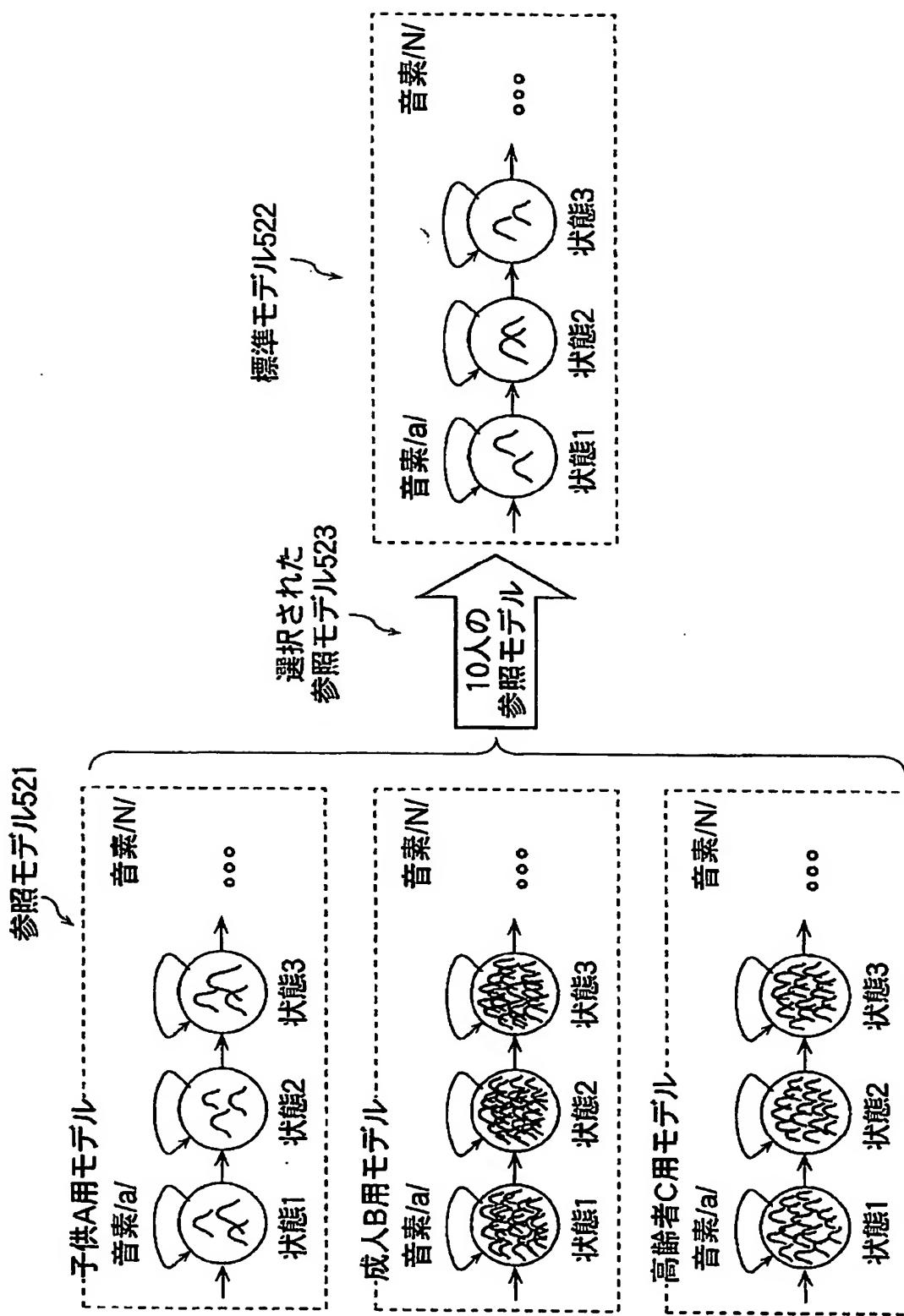
【図23】



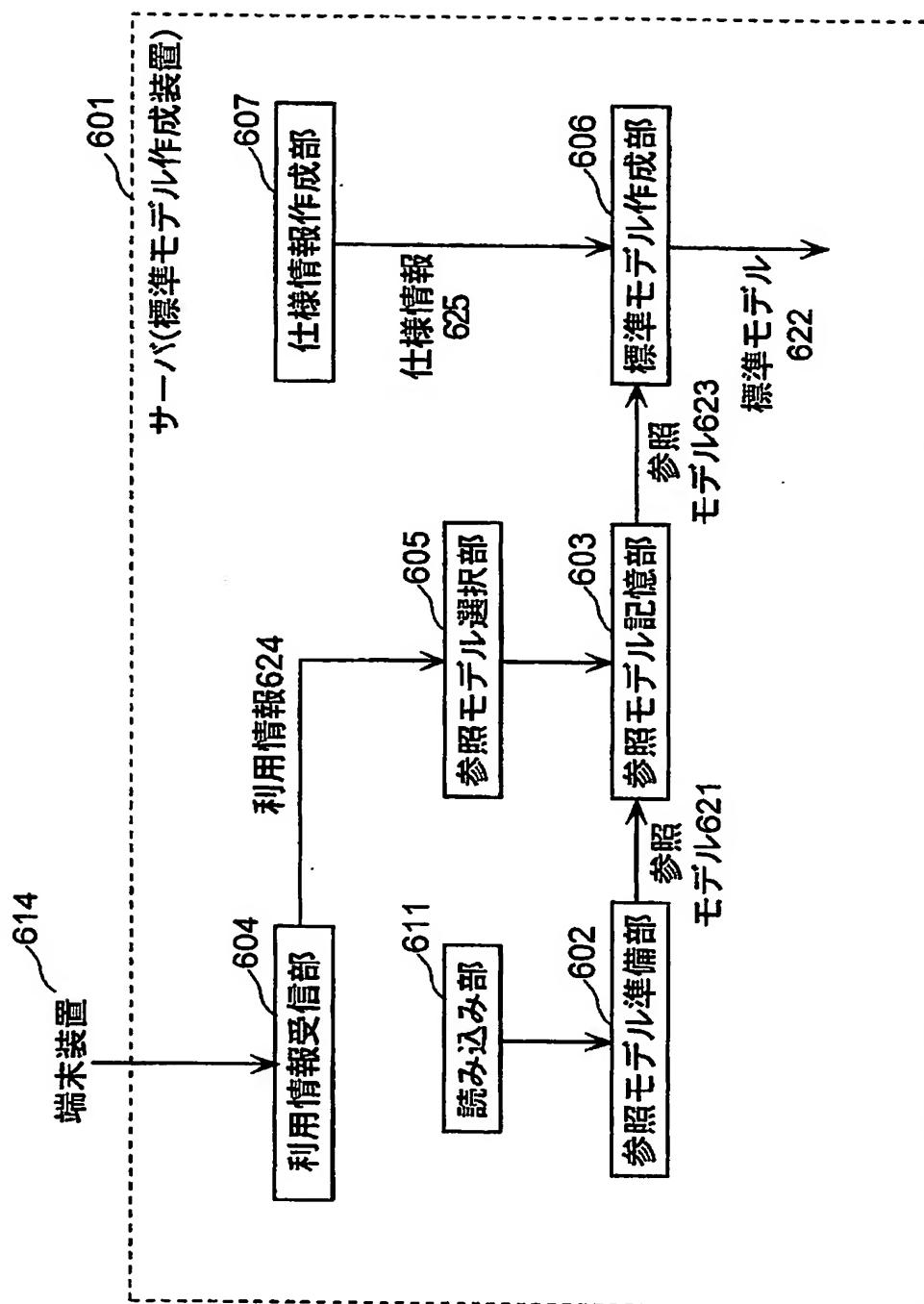
【図24】



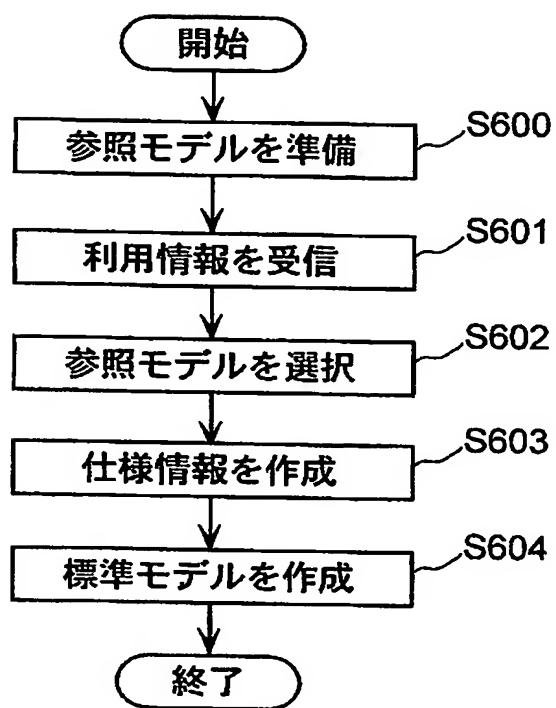
【図25】



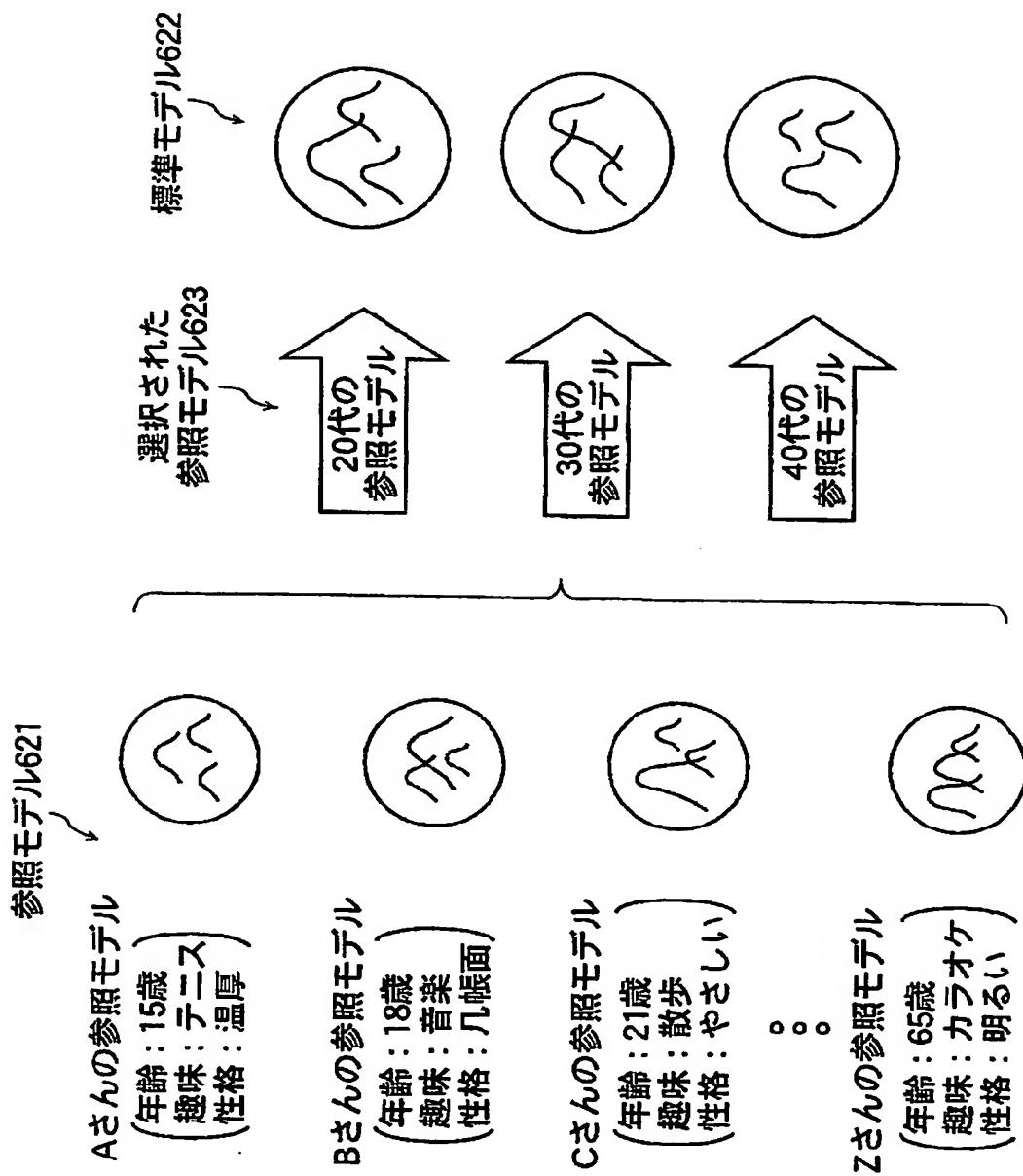
【図26】



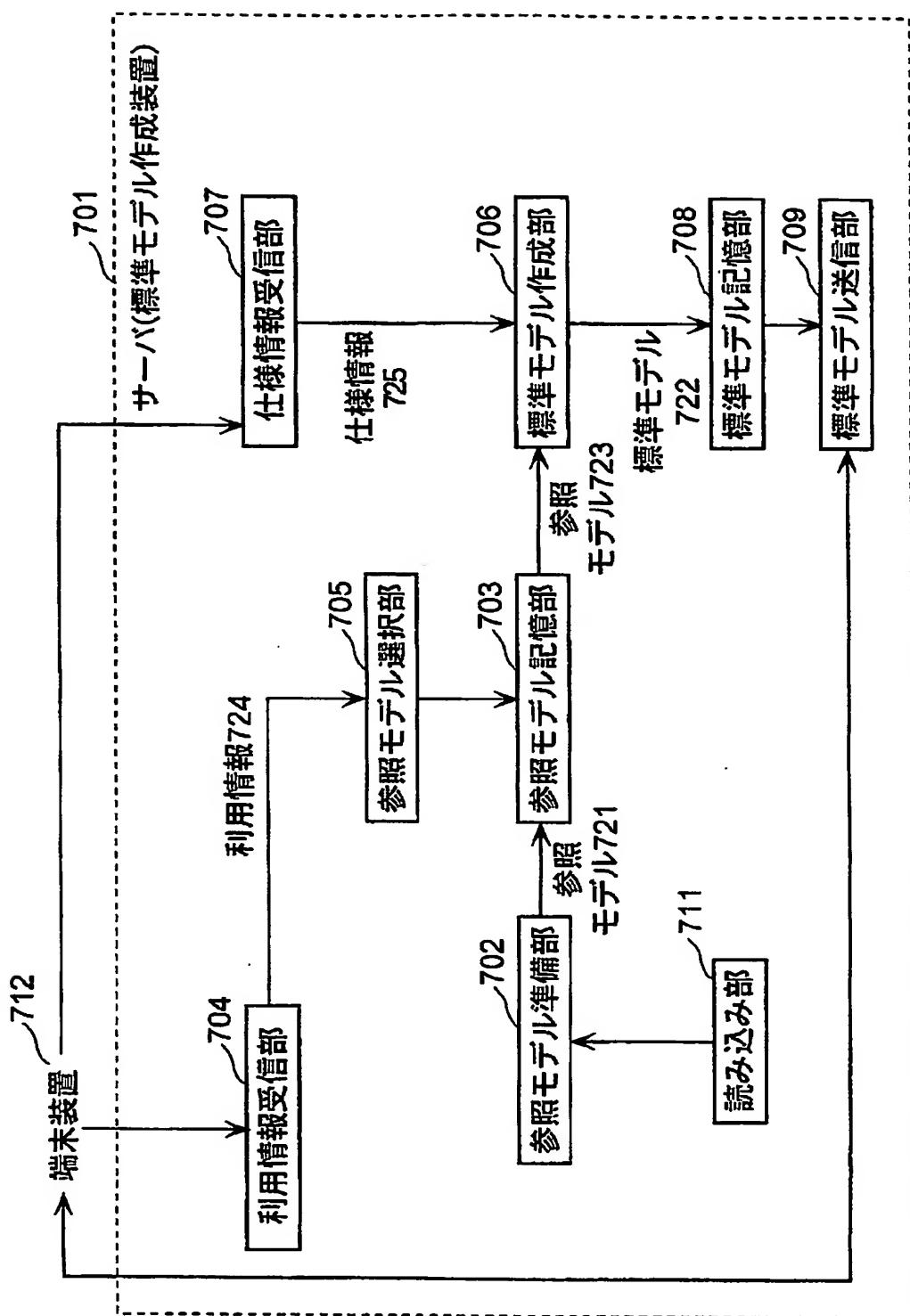
【図27】



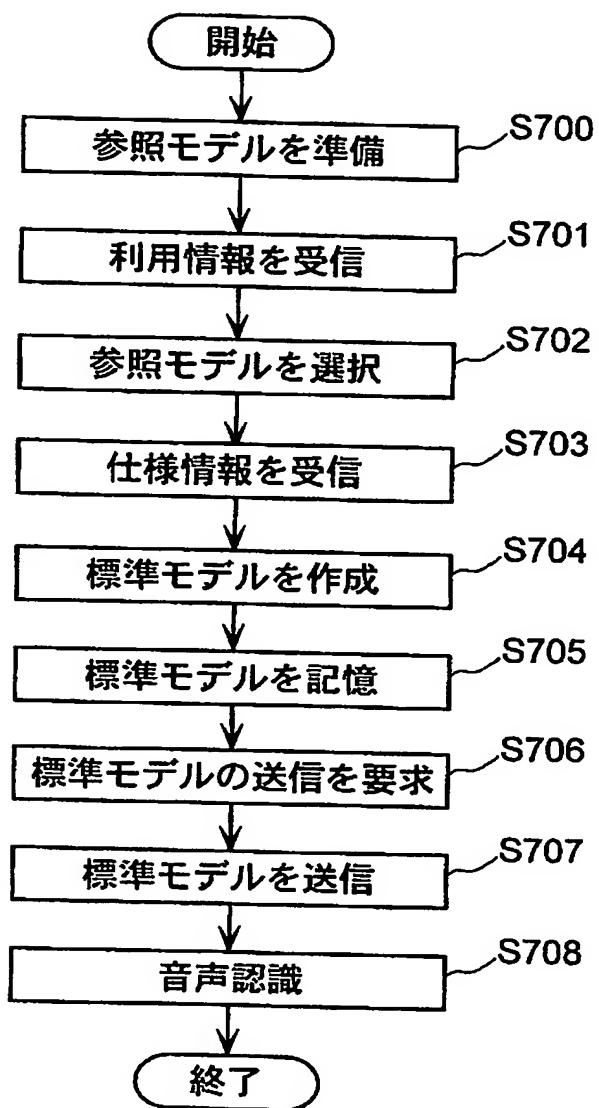
【図28】



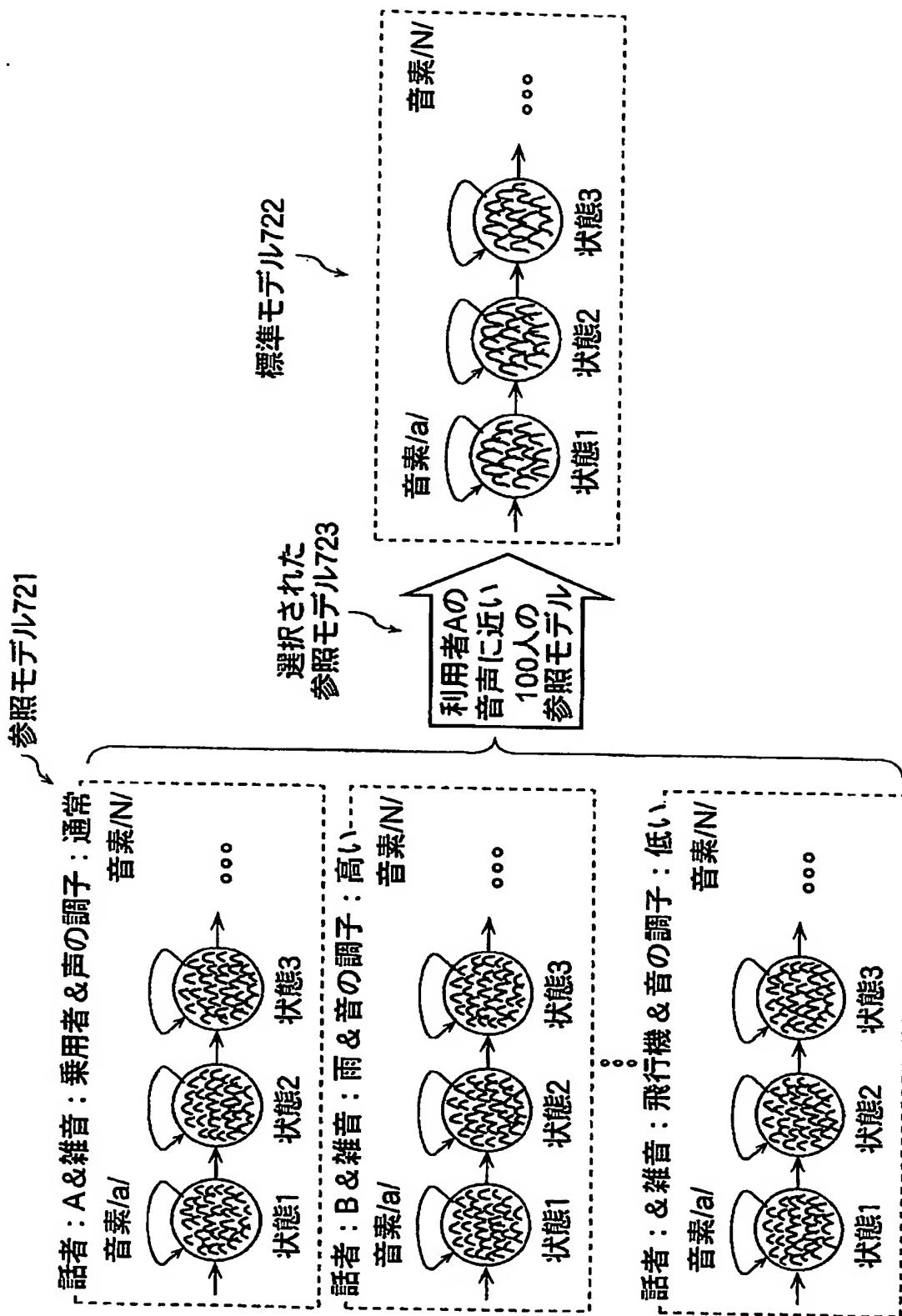
[図29]



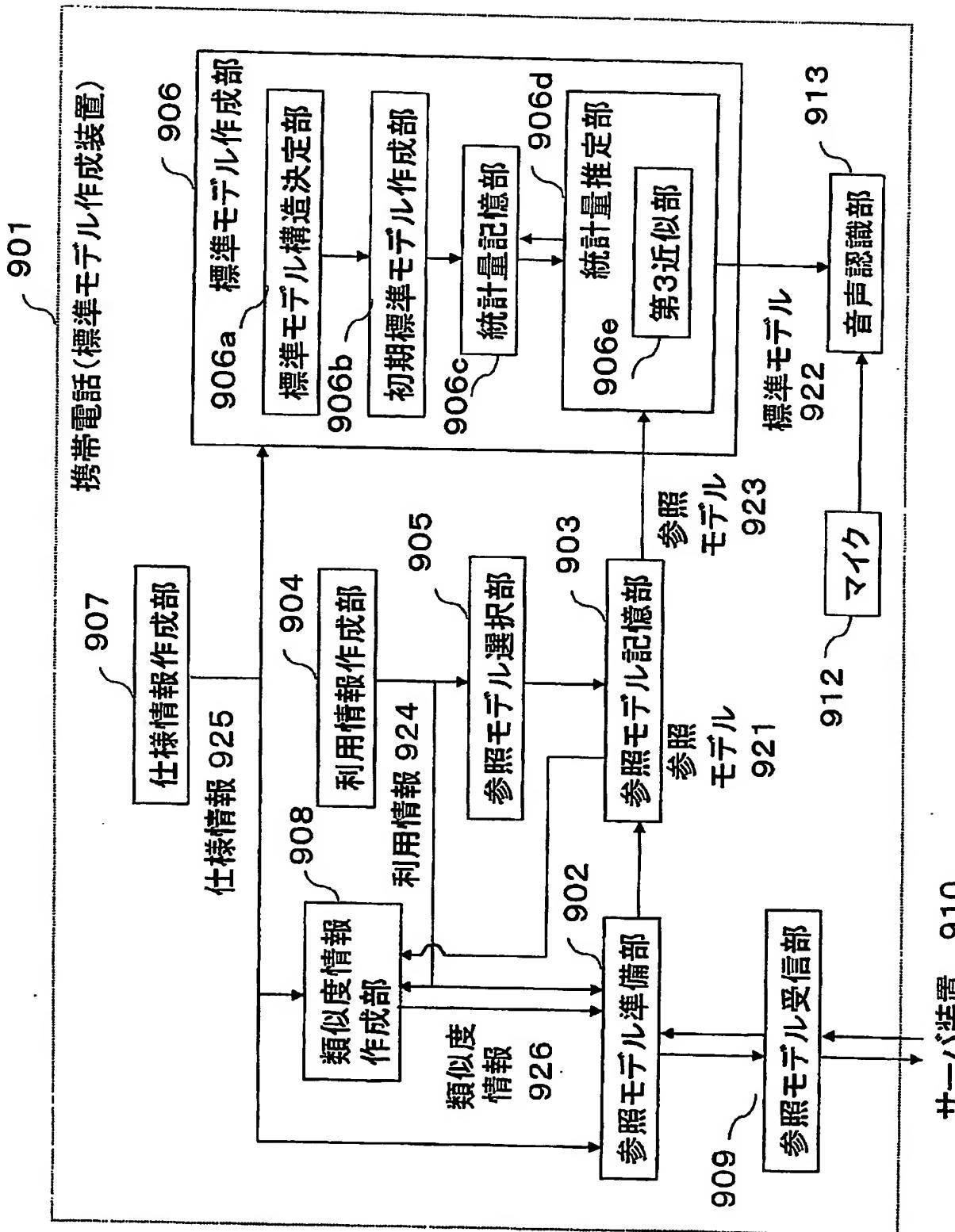
【図30】



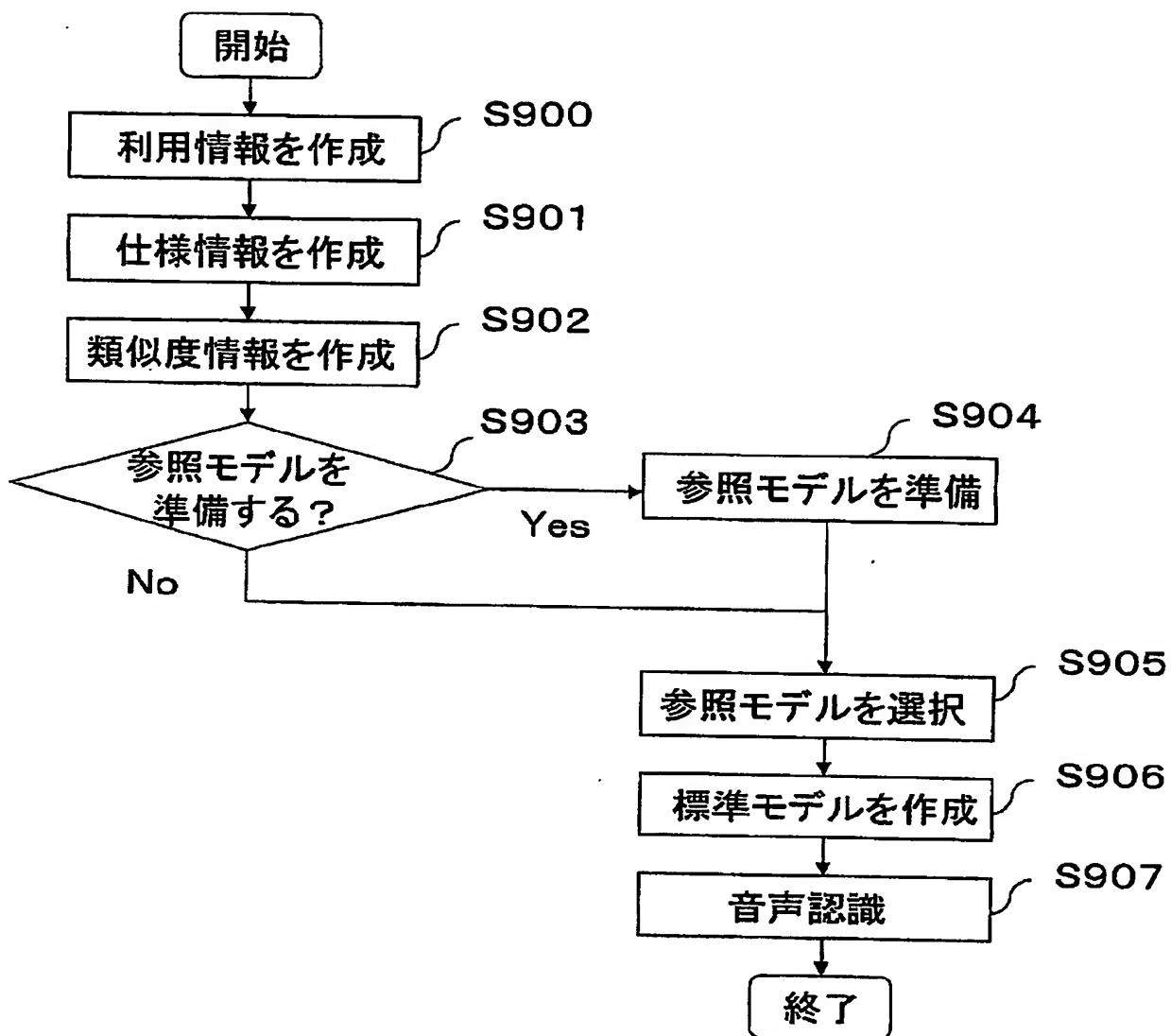
【図31】



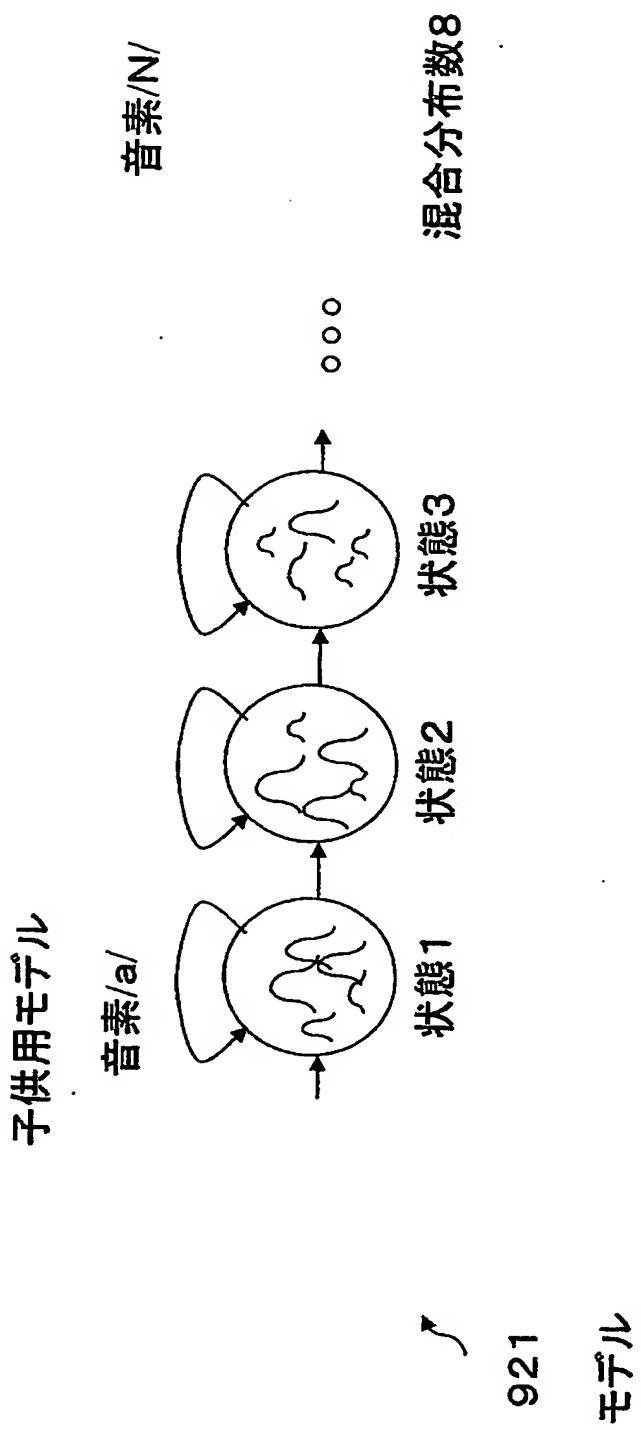
【図32】



【図33】

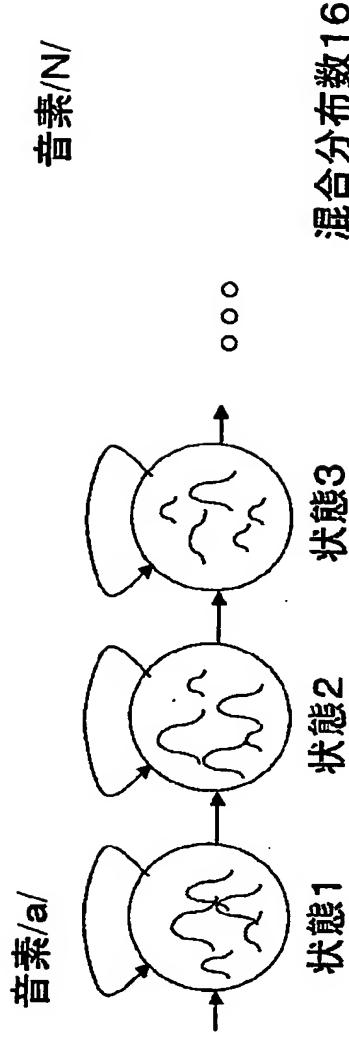


【図34】

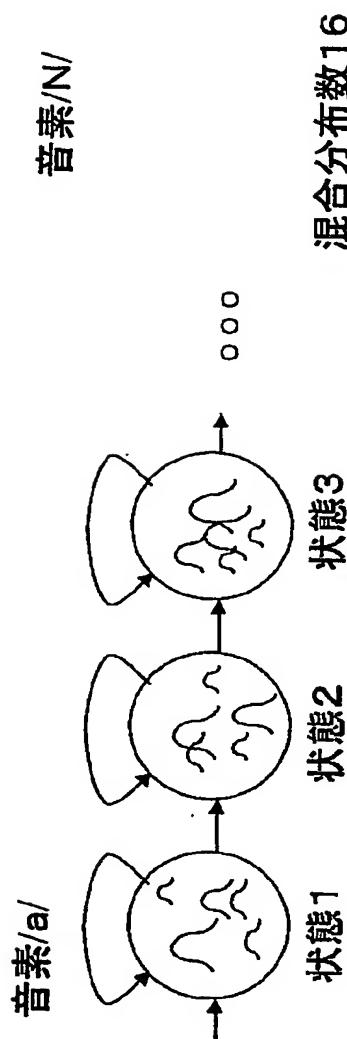


【図35】

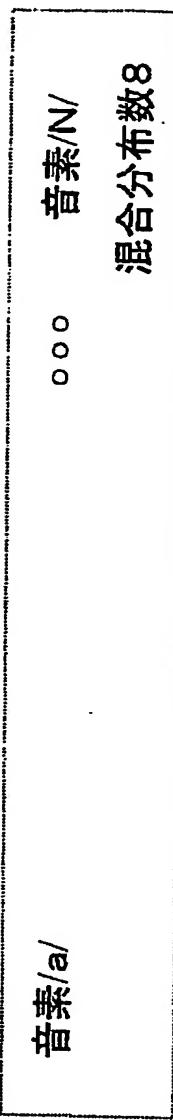
成人女性用モデル



成人男性用モデル

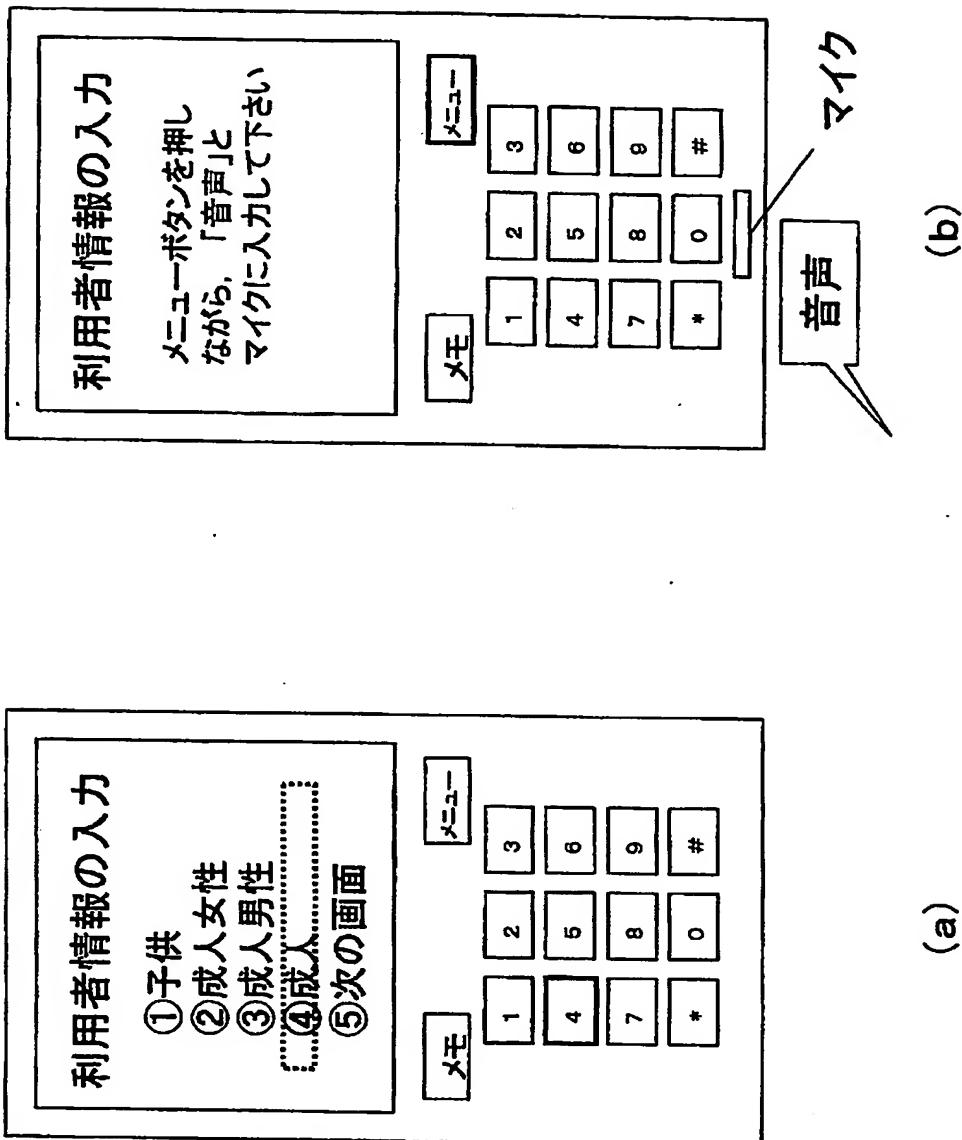


子供用モデル

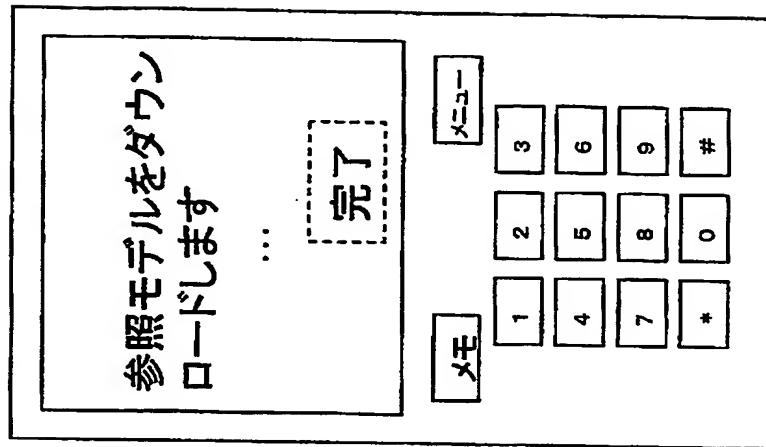


921
モデル

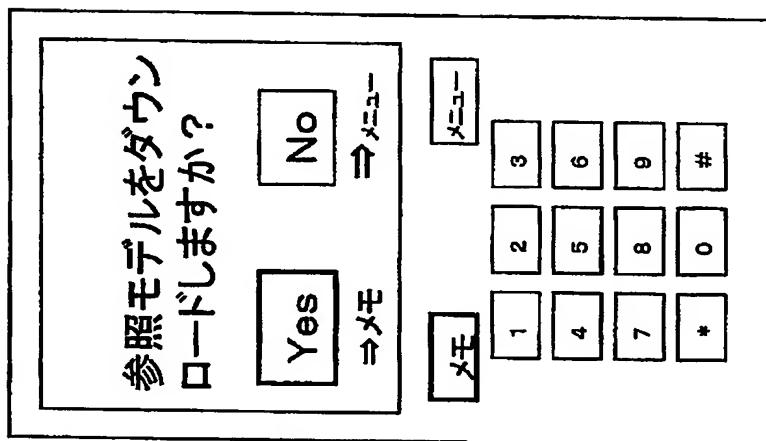
【図36】



【図37】

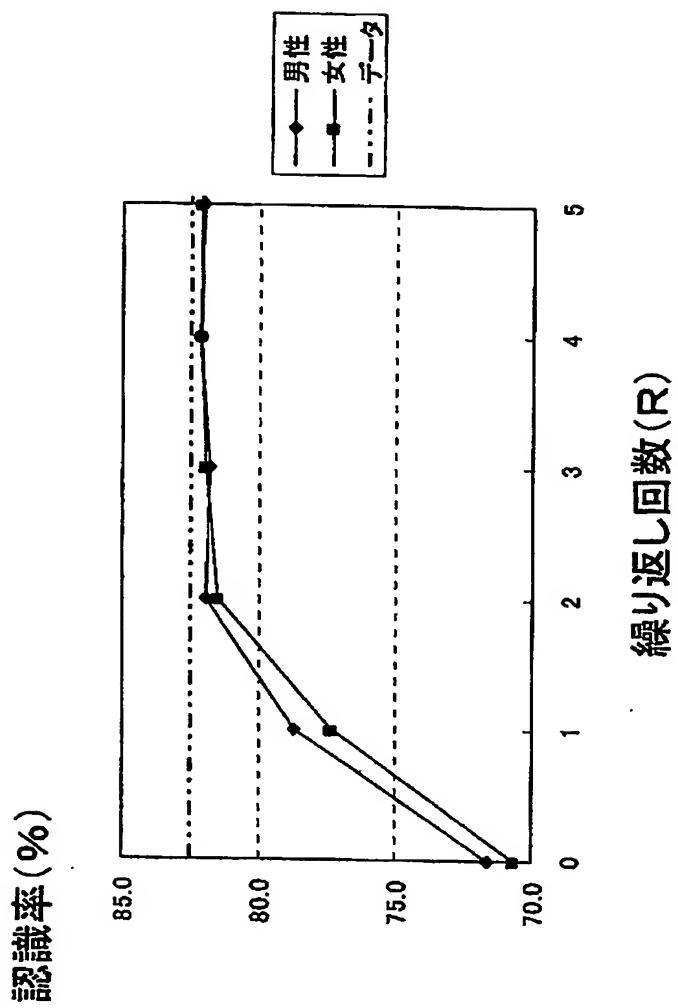


(b)

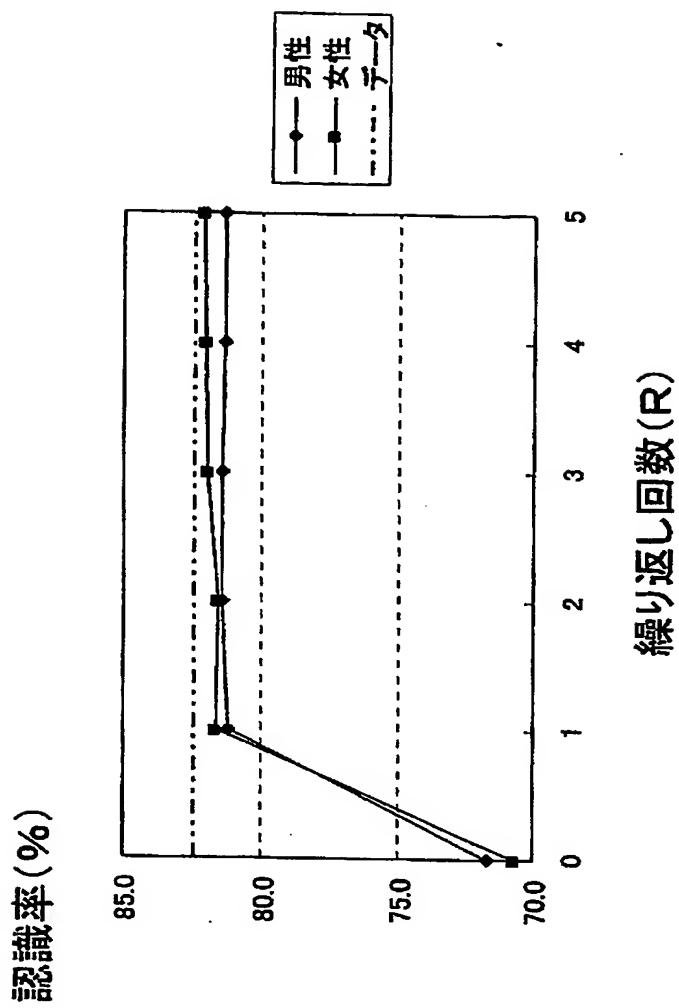


(a)

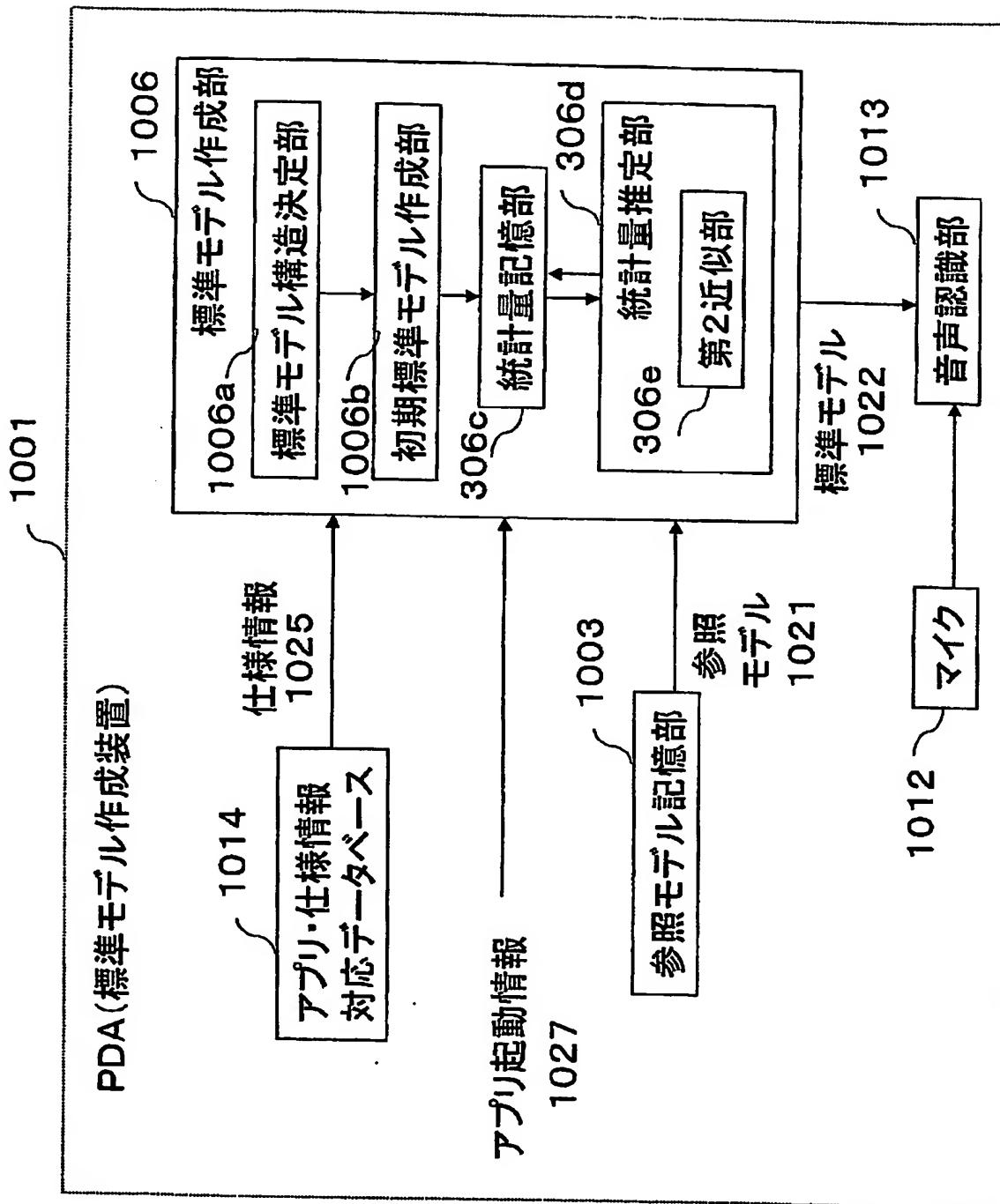
【図38】



【図39】



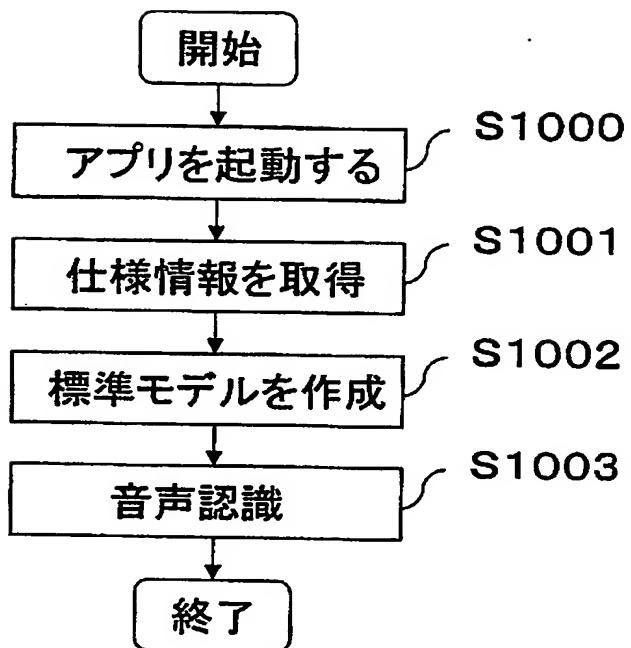
【図40】



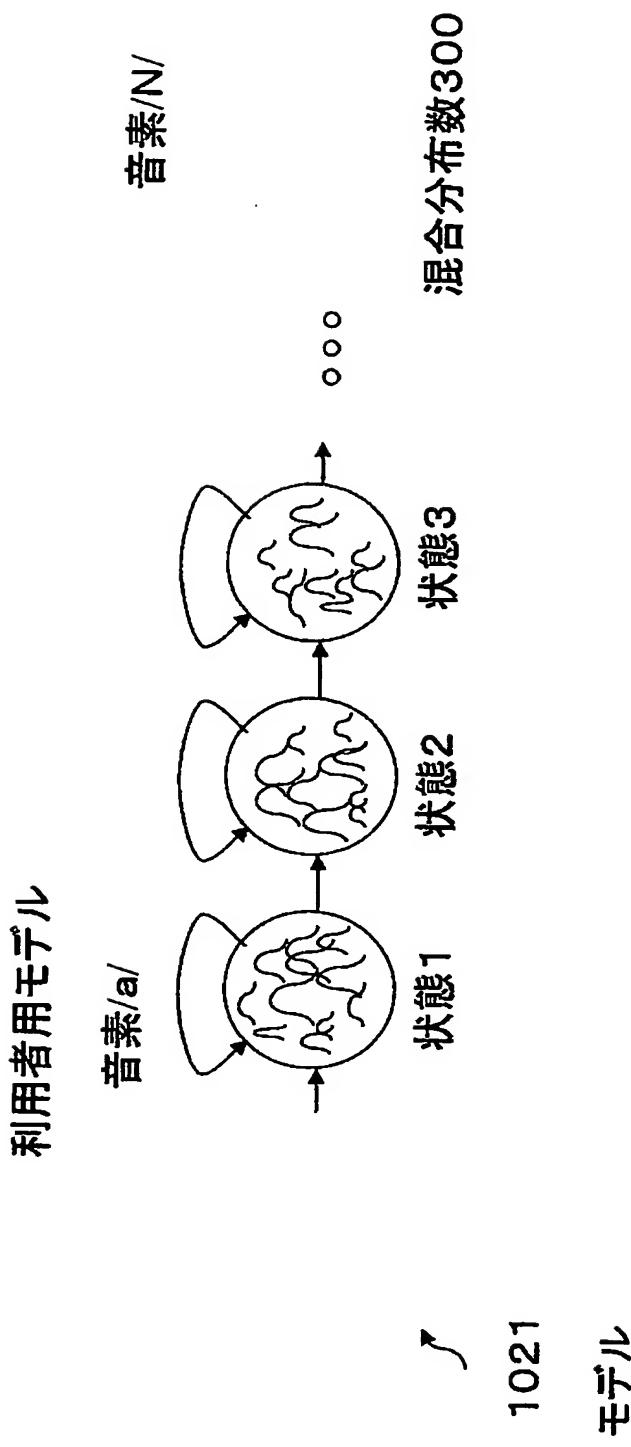
【図41】

アプリケーション		仕様情報
ID	名前	
1	ゲームA	混合分布数3
2	ゲームB	混合分布数5
3	株取引	混合分布数126
4	テレビのリモコン	混合分布数5
5	翻訳	混合分布数64

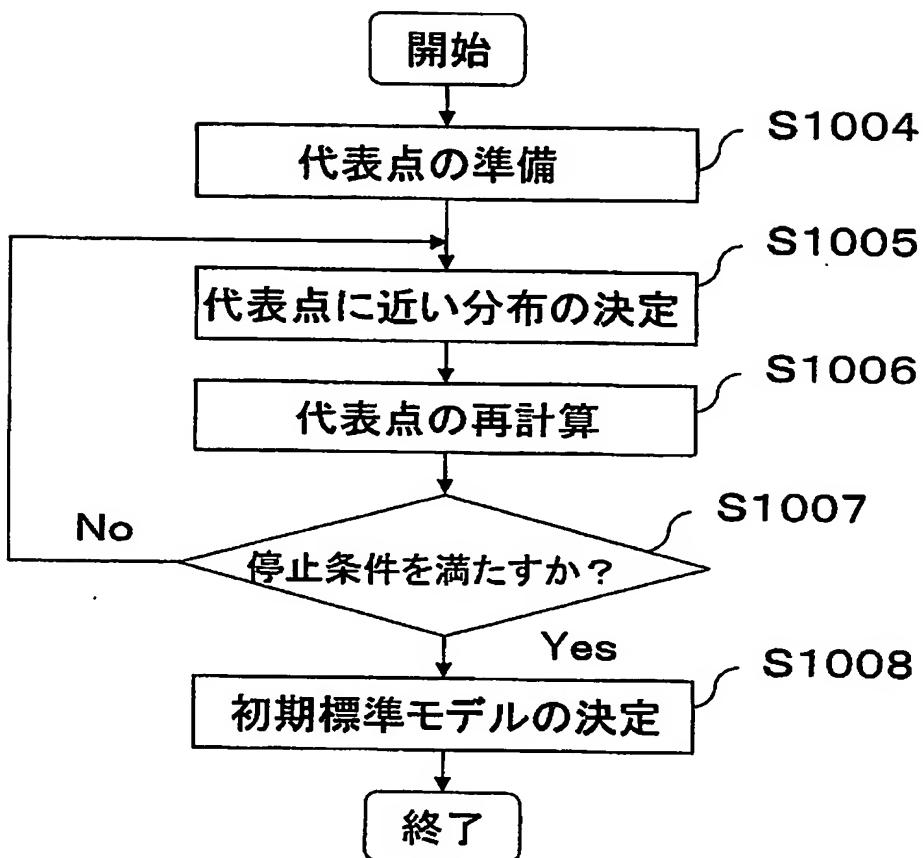
【図42】



【図43】

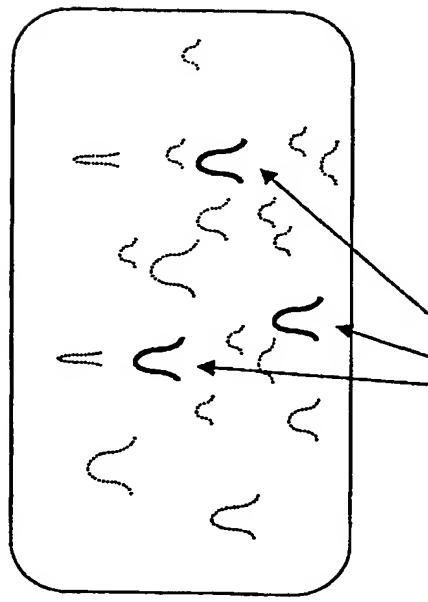


【図44】

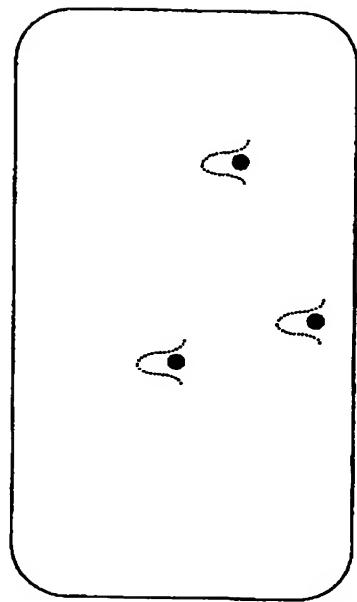


【図45】

参照モデル(300個の分布)

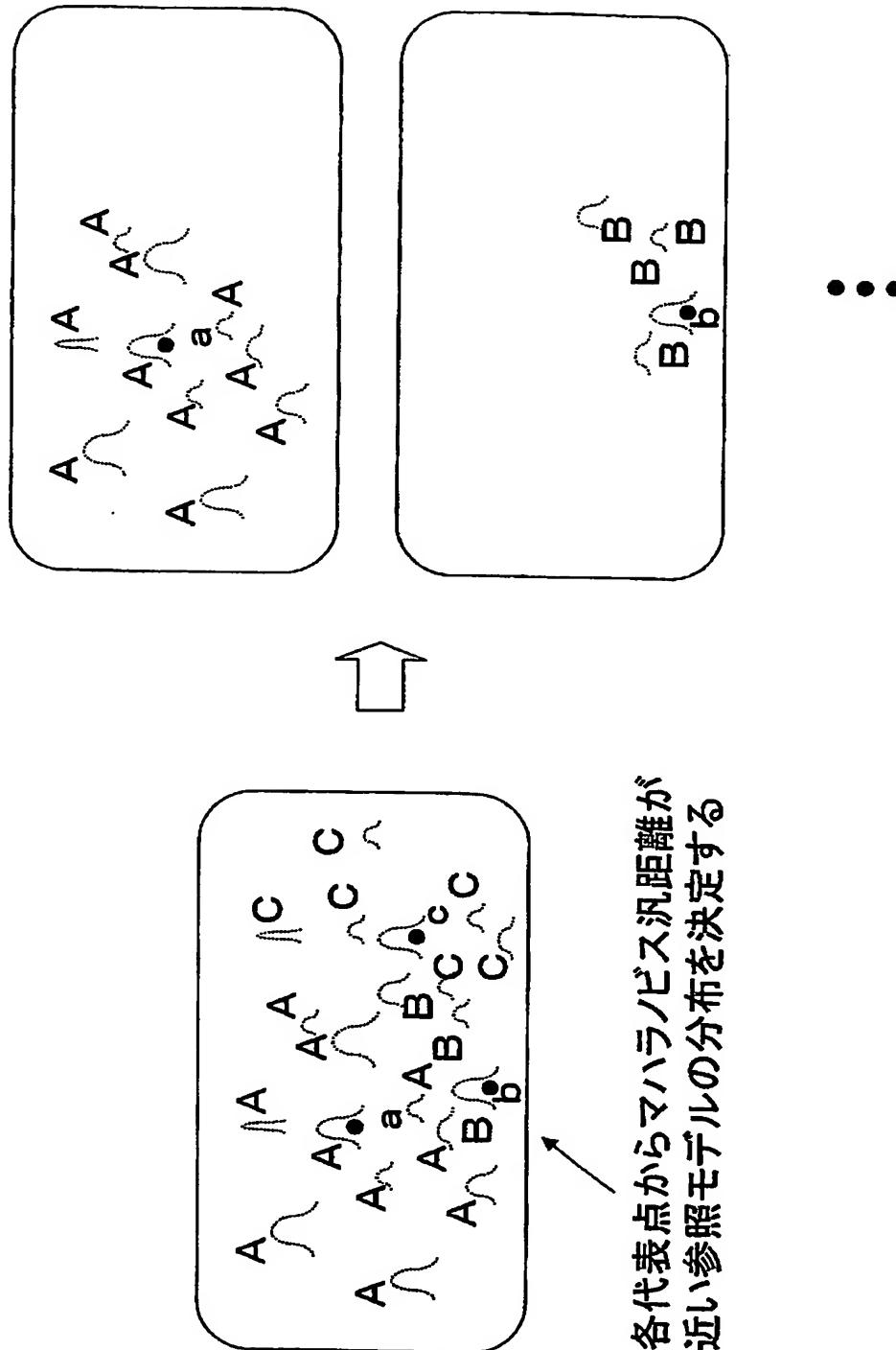


126個(標準モデルの分布数)を選択する

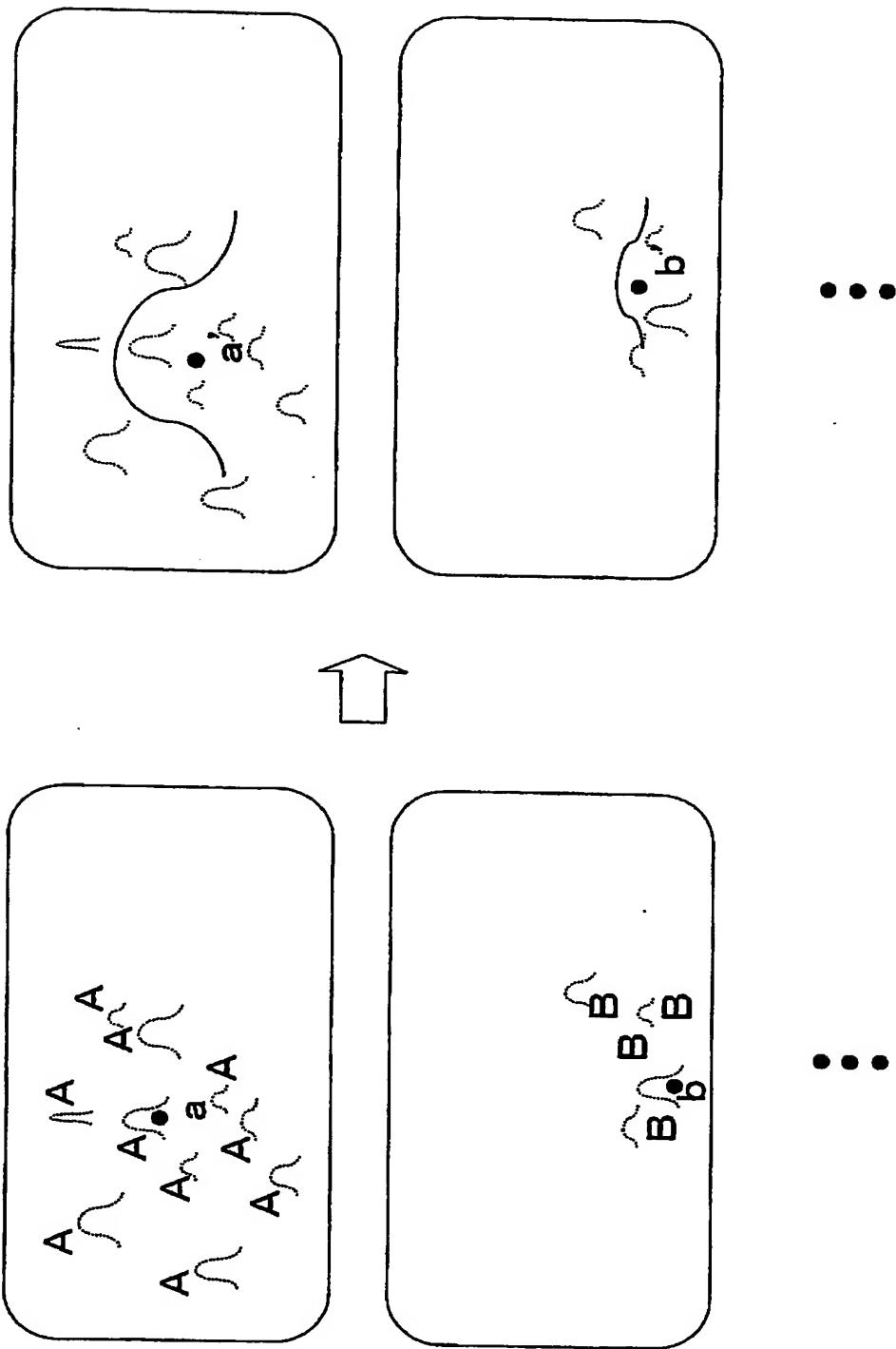


選択された126個の分布の平均値を
代表点とする

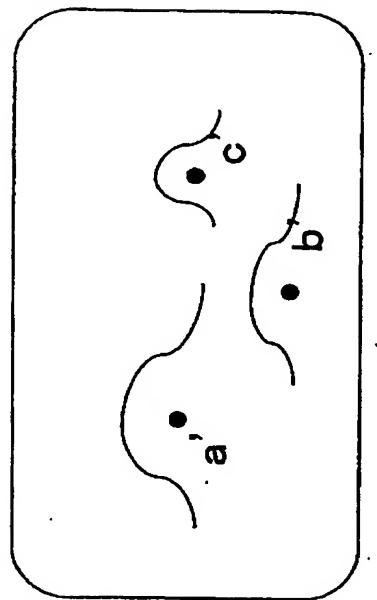
【図46】



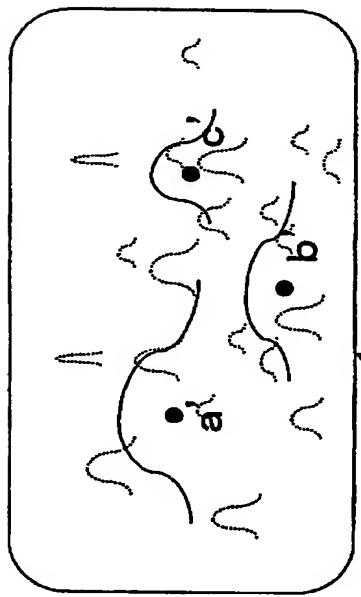
【図47】



【図48】

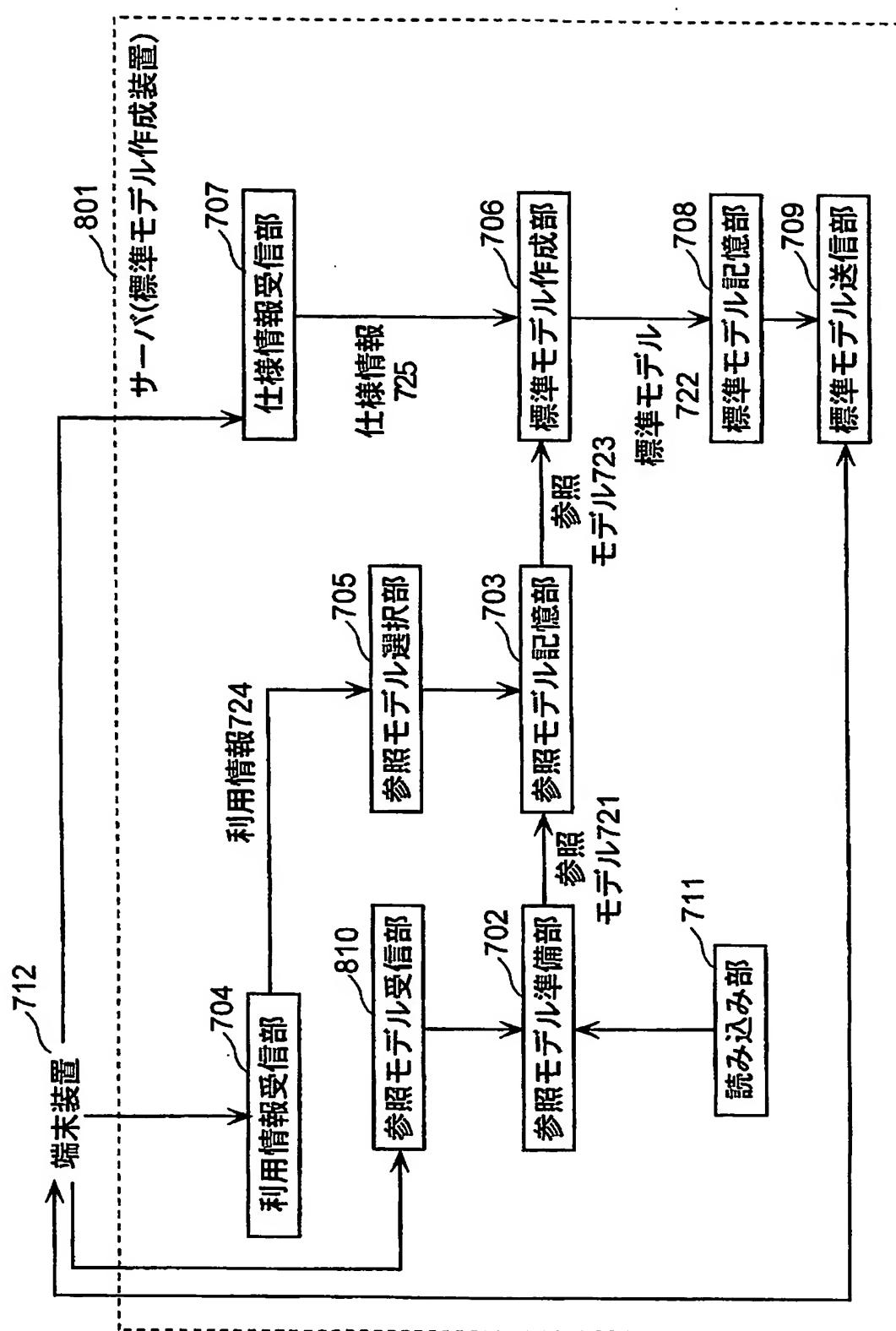


初期標準モデル(126個の分布)を
決定する

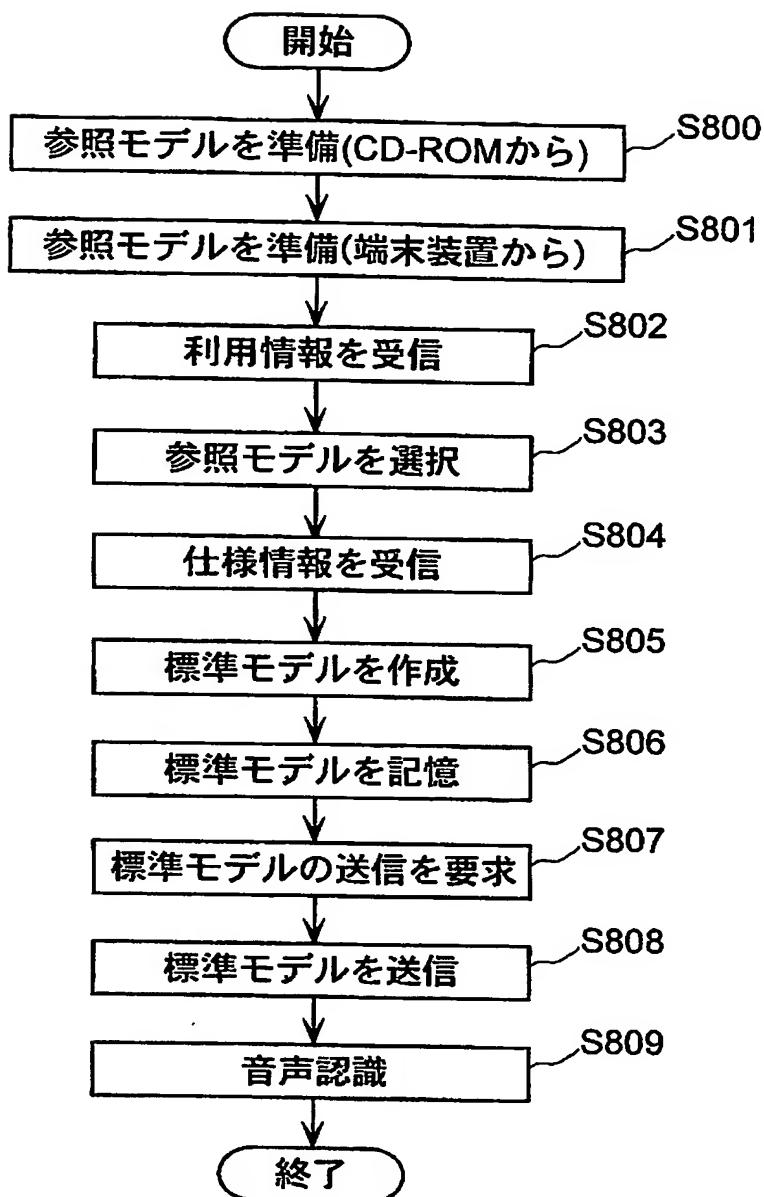


各代表点からマハラノビス汎距離が
近い参照モデルの分布を用いて
新たな分布を作成する

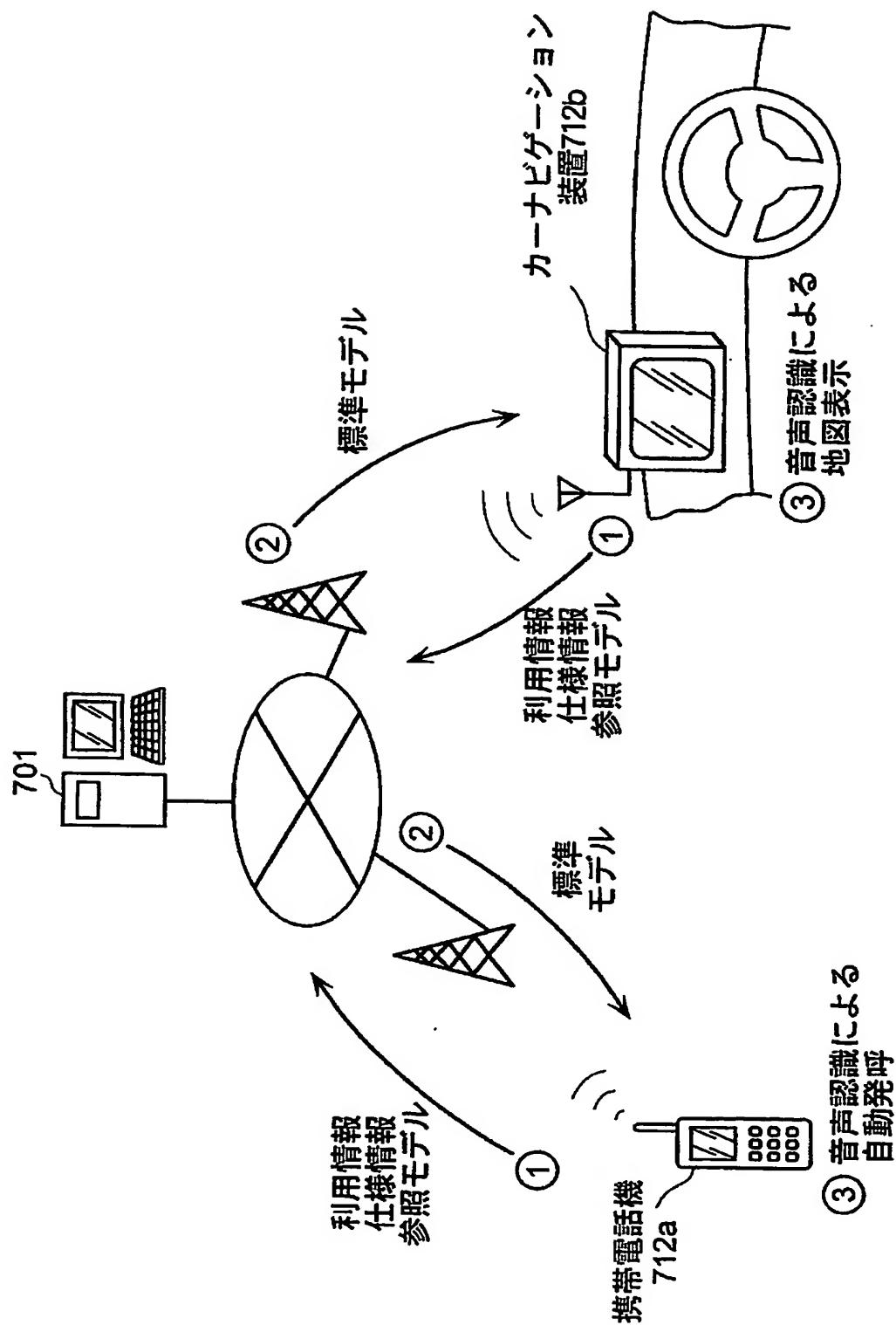
【図49】



【図 50】



【図51】



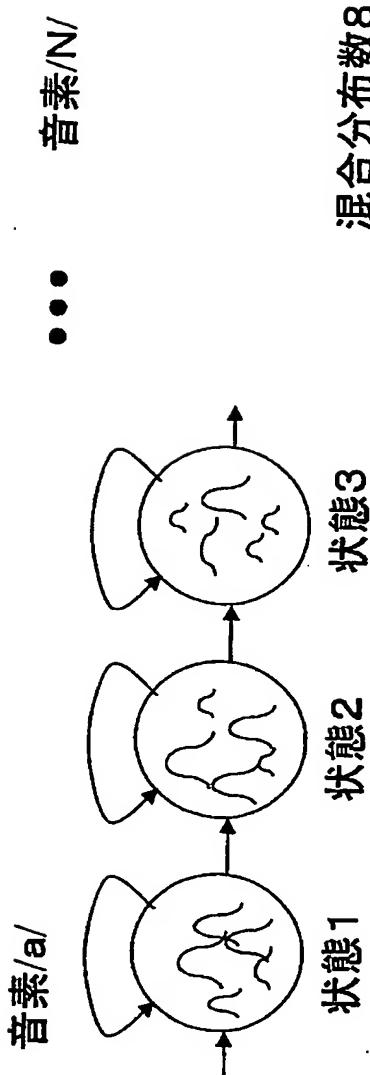
【図52】

クラスID・初期標準モデル・参照モデル対応表

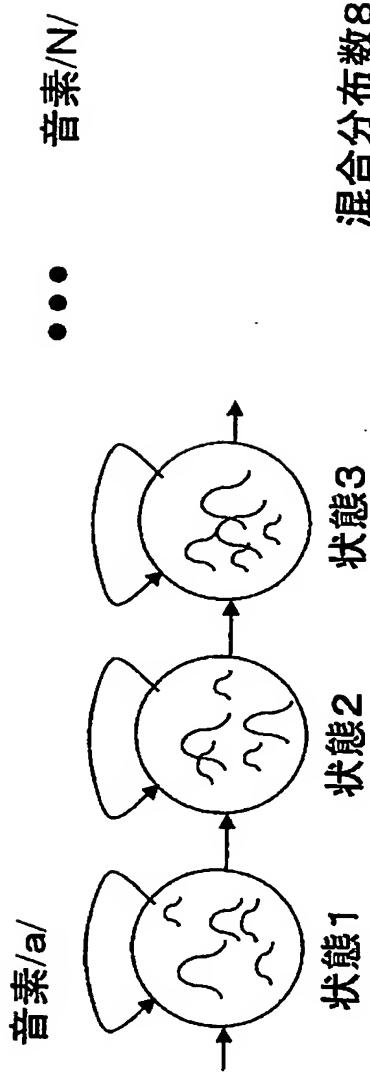
クラスID	初期標準モデル	参照モデル
8A	初期標準モデル8A	参照モデル8AA 参照モデル8AB 参照モデル8AC ⋮ 参照モデル8AZ
⋮	⋮	⋮
64Z	初期標準モデル64Z	参照モデル64ZA 参照モデル64ZB 参照モデル64ZC ⋮ 参照モデル64ZZ

【図53】

参照モデル8AA (クラスID:8A)

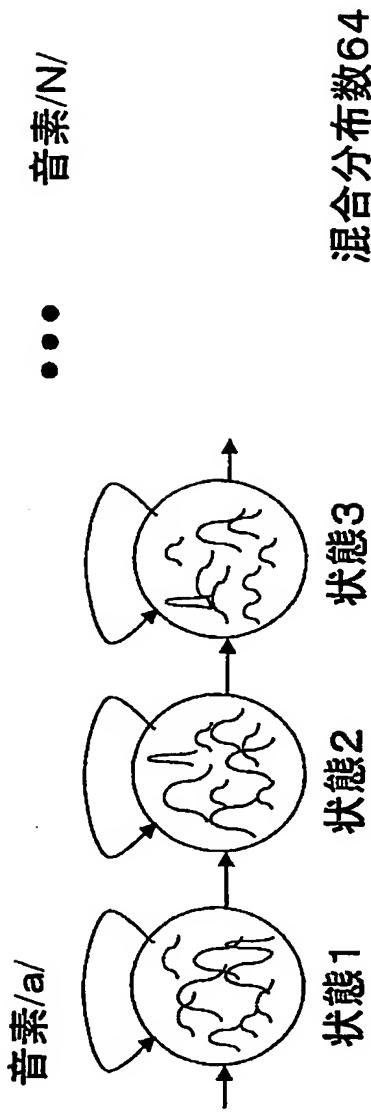
参照モデル
↗

参照モデル8AZ (クラスID:8A)

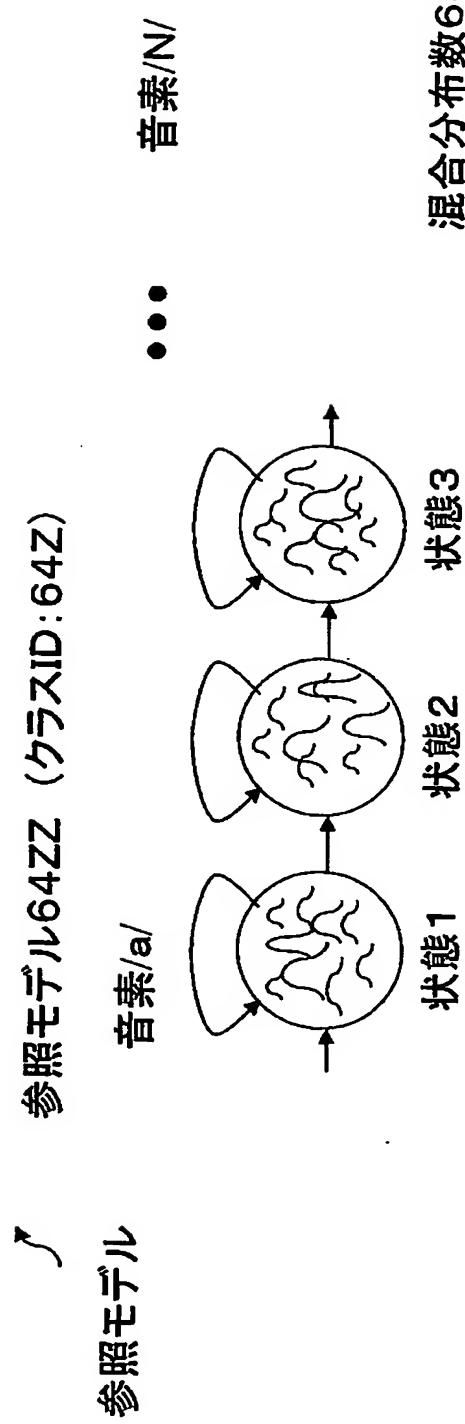


【図54】

参照モデル64ZA (クラスID:64Z)

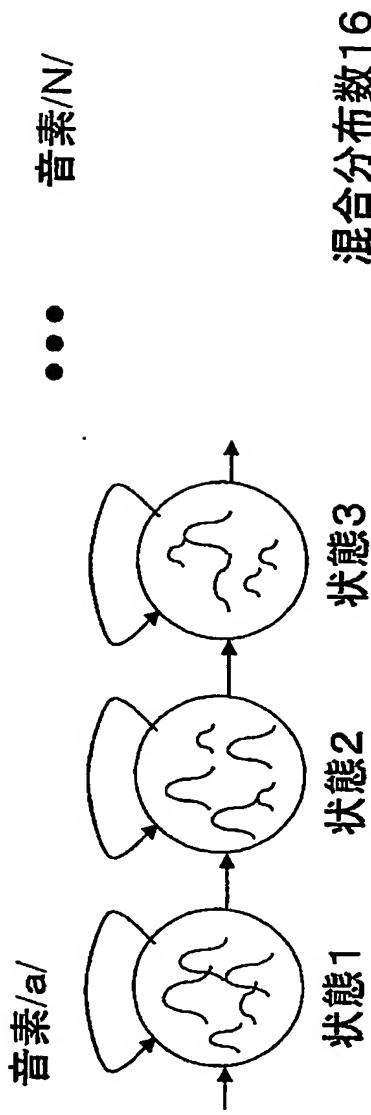


参照モデル64ZZ (クラスID:64Z)

↗ 参照モデル
参照モデル

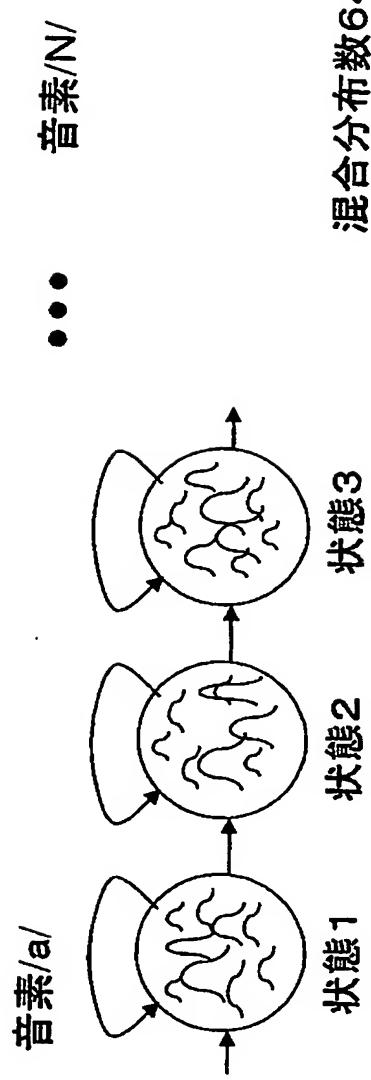
【図55】

初期標準モデル8A (クラスID:8A)

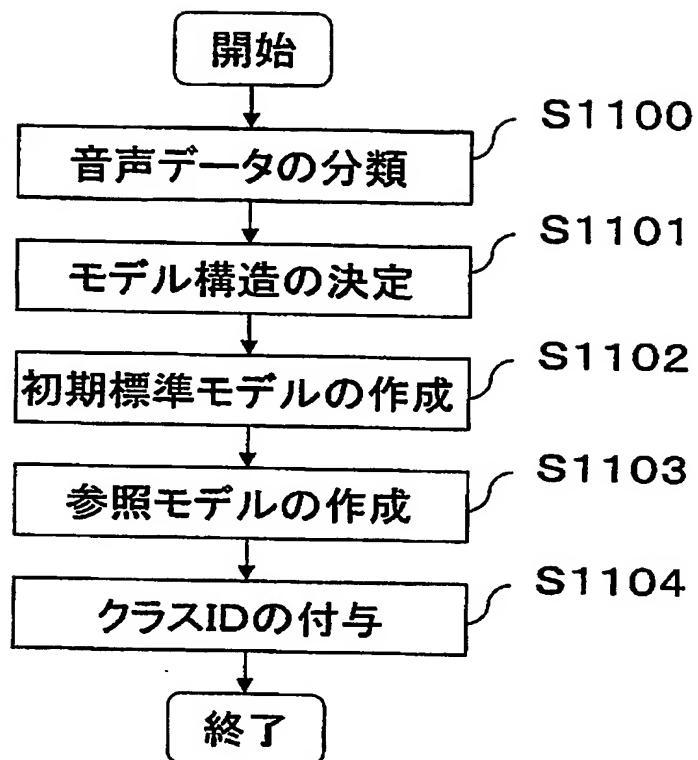


初期標準モデル64Z (クラスID:64Z)

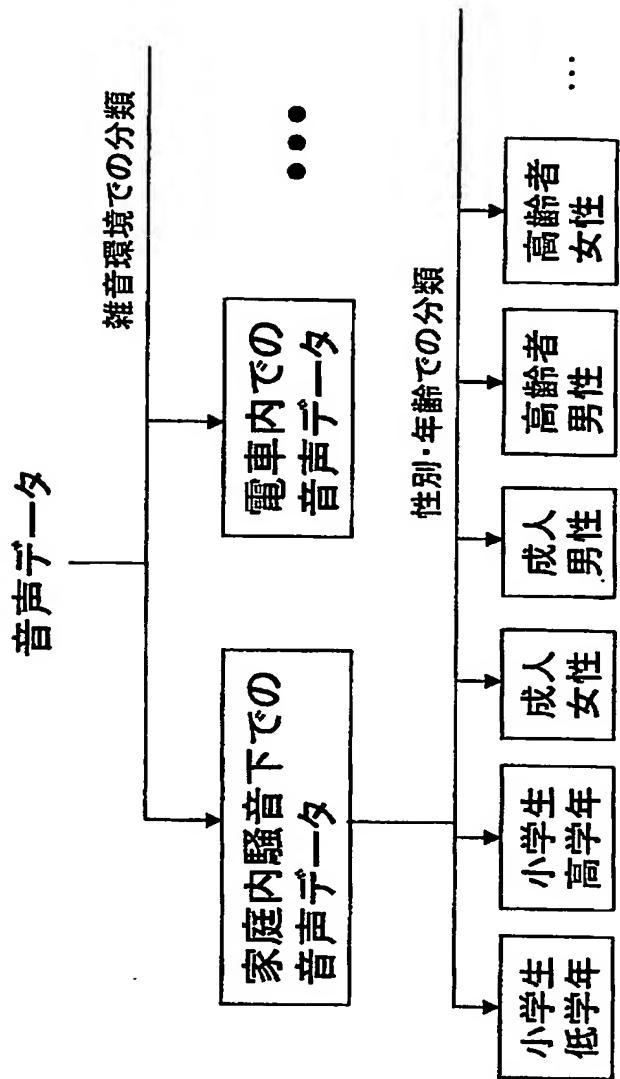
初期標準モデル
初期標準モデル



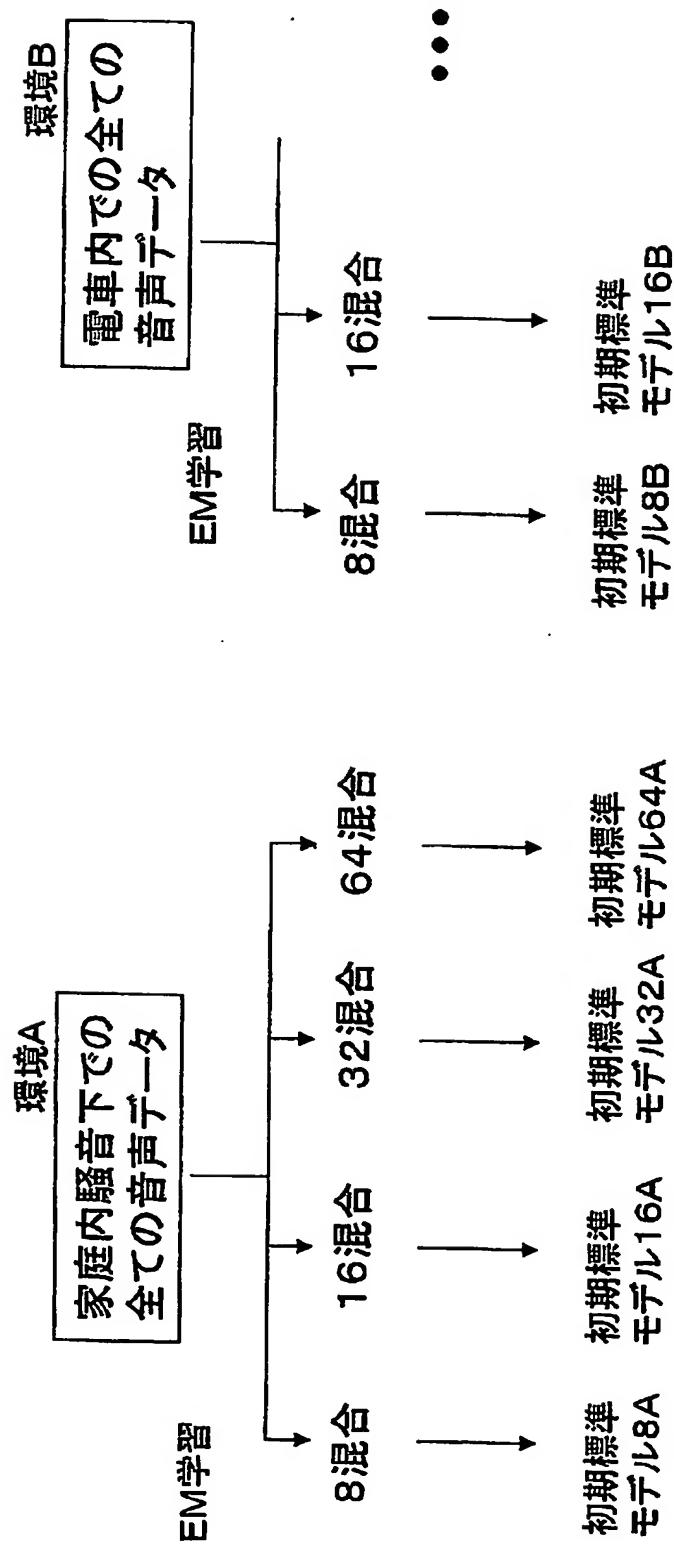
【図56】



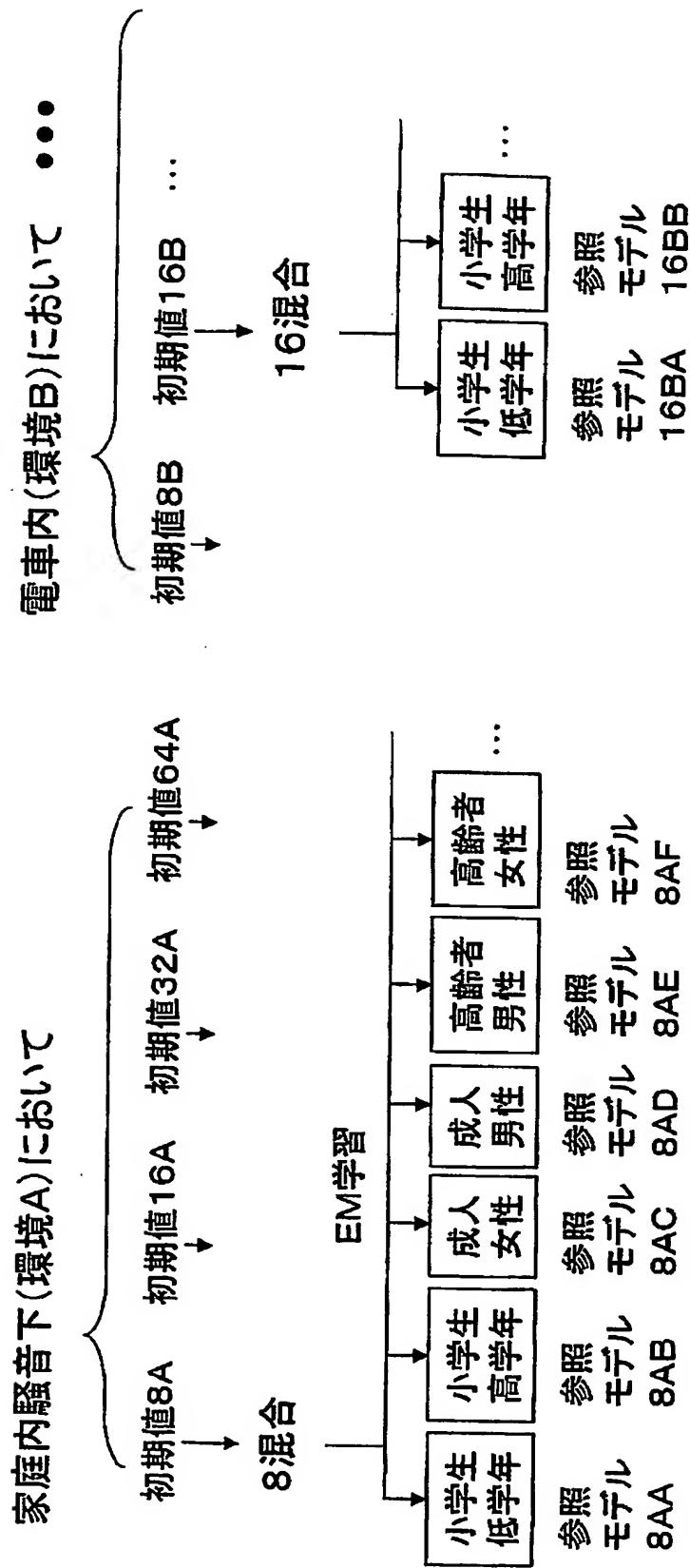
【図57】



【図 5.8】



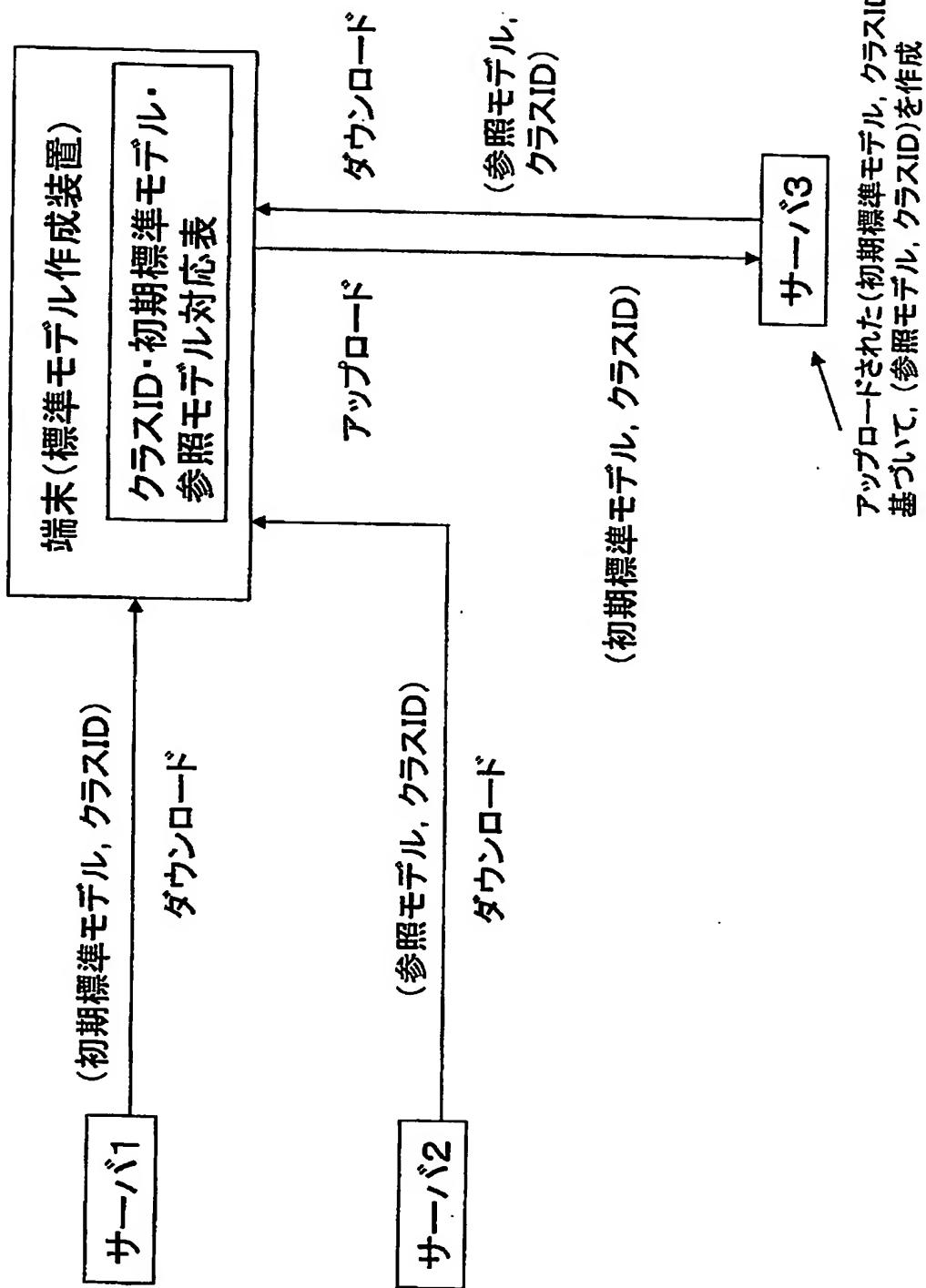
【図59】



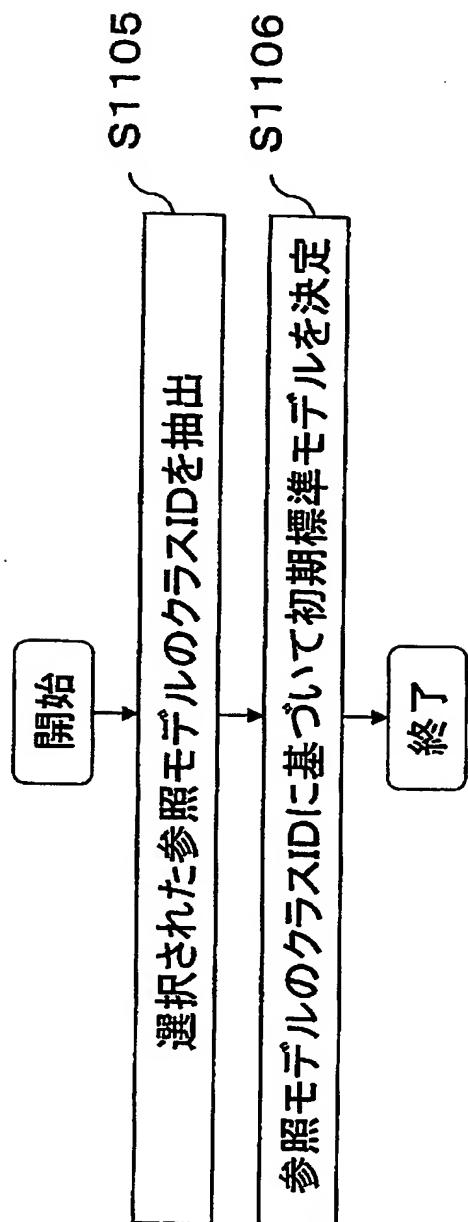
【図60】

クラスID	初期標準モデル	参照モデル	備考（参照モデルの特長）
8A	初期標準モデル8A	参照モデル8AA 参照モデル8AB 参照モデル8AC ⋮	家庭内騒音、8混合、小学校低学年 家庭内騒音、8混合、小学校高学年 家庭内騒音、8混合、成人女性 ⋮
16A	初期標準モデル16A	参照モデル16AA 参照モデル16AB 参照モデル16AC ⋮	家庭内騒音、16混合、小学校低学年 家庭内騒音、16混合、小学校高学年 家庭内騒音、16混合、成人女性 ⋮
⋮	⋮	⋮	⋮
64B	初期標準モデル64B	参照モデル64BA 参照モデル64BB 参照モデル64BC ⋮	電車内、64混合、小学校低学年 電車内、64混合、小学校高学年 電車内、64混合、成人女性 ⋮

【図61】



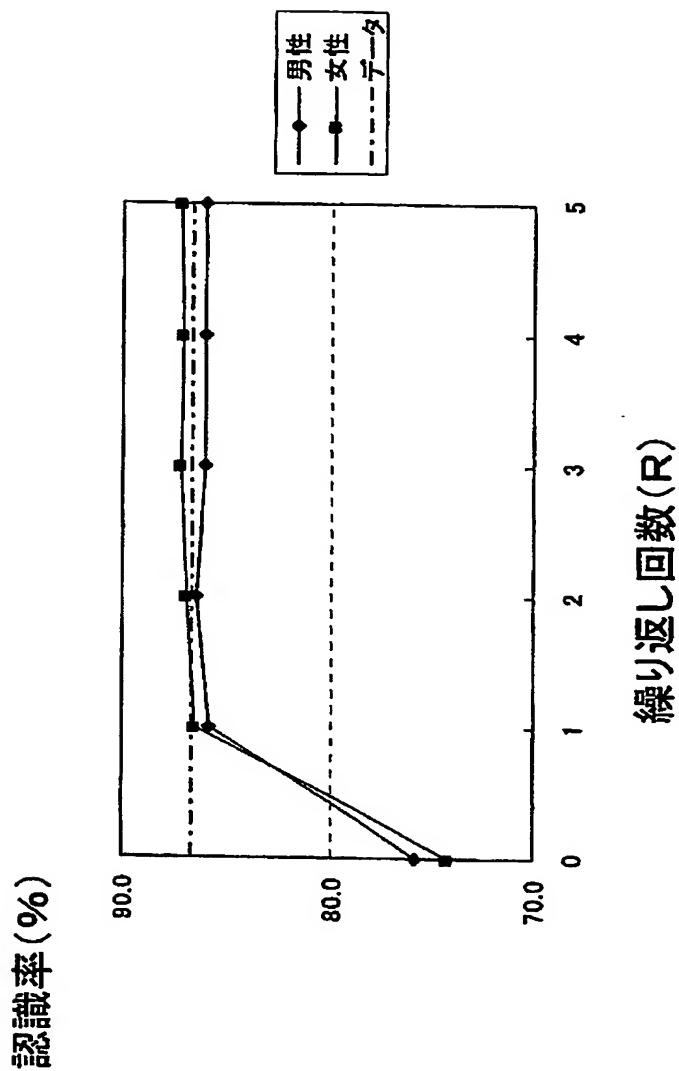
【図62】



【図63】

選択された参照モデル	クラスID
参照モデル8AA	8A
参照モデル16AA	16A
参照モデル16AB	16A
参照モデル16AC	16A
参照モデル16BA	16B
参照モデル64BA	64B

【図64】



【書類名】要約書

【要約】

【課題】 隠れマルコフモデル、ベイズ理論、線形判別分析などの確率モデルによる音声認識、文字認識、画像認識などのパターン認識、ペイジアンネットなどの確率モデルによる意図理解、確率モデルによるデータマイニングなどに用いる高精度な標準モデルを提供する。

【解決手段】 1 以上の参照モデルを準備する参照モデル準備部 102 と、参照モデル準備部 102 が準備した参照モデル 121 を記憶する参照モデル記憶部 103 と、参照モデル記憶部 103 が記憶している 1 以上の参照モデルに対する確率または尤度を最大化または極大化するように標準モデルの統計量を計算して標準モデル 122 を作成する標準モデル作成部 104 とを備える。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願2003-284489
受付番号	50301276153
書類名	特許願
担当官	第八担当上席 0097
作成日	平成15年 8月 5日

<認定情報・付加情報>

【提出日】	平成15年 7月31日
-------	-------------

特願2003-284489

出願人履歴情報

識別番号 [000005821]

1. 変更年月日 1990年 8月28日

[変更理由] 新規登録

住所 大阪府門真市大字門真1006番地
氏名 松下電器産業株式会社